

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 13:11:21 ; Search time 1906 Seconds

(without alignments)
5725.893 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 375
Sequence: 1 atgaagatattcttgatttc.....atgaagtgtgaagattaa 375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.ov:*
22: em.or:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vl:*
30: em.htg.hum:*
31: em.htg.in:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pln:*
35: em.htg.rtd:*
36: em.htg.mam:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	375	100.0	375	6	AX036541	AX036541 Sequence
2	48.6	13.0	189967	2	AL833802	AL833802 Mus muscu
3	48.6	13.0	206914	10	AL807815	AL807815 Mouse DNA
4	46.4	12.4	40090	3	AC006611	AC006611 Caenorhab
5	46.4	12.4	299670	2	AC006845	AC006845 Caenorhab
6	45.8	12.2	4458	2	AC115614	AC115614 Dictyoste
7	45.8	12.2	49999	6	AX015902	AX015902 Sequence
8	45.8	12.2	49999	6	AX015908	AX015908 Sequence
9	45.8	12.2	119191	9	AP004782	AP004782 Homo sapi
10	45.8	12.2	228283	10	AF312994	AF312994 Mus muscu
11	45.2	12.1	20796	3	AF216973	AF216973 Drosophi
12	45.2	12.1	87767	2	AC014497	AC014497 Drosophi
13	45.2	12.1	132000	2	AC116976	AC116976 Dictyoste
14	45.2	12.1	167062	2	AC007624	AC007624 Drosophi
15	45.2	12.1	181771	3	AC008340	AC008340 Drosophi
16	45.2	12.1	348873	3	AE003790	AE003790 Drosophi
17	44.8	11.9	26671	3	CER11A5	283122 Caenorhabd
18	44.4	11.8	2320	9	BC002875	BC002875 Homo sapi
19	44.4	11.8	3230	9	BC010457	BC010457 Homo sapi
20	44.4	11.8	86945	2	AC002490	AC002490 Homo sapi
21	44.4	11.8	153477	2	AC006278	AC006278 Plasmodu
22	44.4	11.8	154470	2	AC027820	AC027820 Homo sapi
23	44.4	11.8	195770	2	AC091153	AC091153 Homo sapi
24	44.2	11.8	247175	2	AC129593	AC129593 Mus muscu
25	43.6	11.6	110000	2	PFMAL13P2_1	Continuation (2 of
26	43.6	11.6	152409	2	PFMAL1P1	AL031744 Plasmodu
27	43.6	11.6	183767	2	AL807765	AL807765 Mus muscu
28	43.6	11.6	234295	2	AL714007	AL714007 Mus muscu
29	43.4	11.6	98734	2	PFMAL1P2	AL031745 Plasmodu
30	43.4	11.6	143716	2	AL627248	AL627248 Dario rer
31	43.4	11.6	168083	9	AC092376	AC092376 Homo sapi
32	43.2	11.5	977	11	CNS061DX	AL400171 77 end of
33	43.2	11.5	139665	2	AP000865	AP000865 Homo sapi
34	43	11.5	110000	2	PFMAL1P1	Continuation (2 of
35	43	11.5	162261	2	AC119001	AC119001 Rattus no
36	42.8	11.4	142439	9	AC008749	AC008749 Homo sapi
37	42.6	11.4	59762	8	AB023032	AB023032 Arabidops
38	42.6	11.4	173016	2	AC019123	AC019123 Homo sapi
39	42.6	11.4	262428	2	AC096326	AC096326 Rattus no
40	42.4	11.3	2783	3	AF093702	AF093702 Plasmodu
41	42.4	11.3	49144	2	AC131380	AC131380 Strongylo
42	42.4	11.3	115758	9	AC104634	AC104634 Homo sapi
43	42.4	11.3	198516	2	AC102659	AC102659 Mus muscu
44	42.4	11.3	202521	2	AL773546	AL773546 Mus muscu
45	42.4	11.3	234112	3	PFMAL1P2	AL035475 Plasmodu

ALIGNMENTS

RESULT 1
AX036541
LOCUS AX036541
DEFINITION Sequence 1 from Patent W00056885.
ACCESSION AX036541
VERSION AX036541.1 GI:11226151
KEYWORDS
SOURCE medicinal leech.
ORGANISM Hirudo medicinalis
Eukaryota; Metazoa; Annelida; Clitellata; Hirudinea; Hirudinae;
Arynchobdellida; Hirudiniiformes; Hirudinae; Hirudo.
REFERENCE
1 (bases 1 to 375)
Hemberger,J., Scheuble,B., Strittmatter,W., Hofmann,U., Fotev,Z.
and Gnessow,D.
TITLE Protein for blocking platelet adhesion

JOURNAL Patent: WO 0056885-A 1 28-SEP-2000;
HEMBERGER JUEGEN (DE) ; MERCK PATENT GMBH (DE) ; SCHEUBLE BERNHARD
(DE) ; STRITTMATTER WOLFGANG (DE) ; HOEMANN UWE (DE) ; FOTEV ZISI
(DE) ; GUESSON DETLEF (DE)
FEATURES Location/Qualifiers

```

/organism="Hirudo medicinalis"
/db_xref="taxon:6421"
<64. .375
CDS

```

BASE COUNT	123 a	62 c	85 g	105 t
ORIGIN				

Query Match	100.0%;	Score 375;	DB 6;	Length 375;
Best Local Similarity	100.0%;	Pred. No. 6.6e-86;		
Matches 375;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	ATGAAGTATTTCTGTGATTTCTCTTCTGCGGAGAGCTGATCTCAACTACTCTG	60
Db	1	ATGAAGTATTTCTGTGATTTCTCTTCTGCGGAGAGCTGATCTCAACTACTCTG	60
QY	61	TCAGAAAGACGTGAGAGATTTGTGACGCTTTTACGCGAACAGAAATATATACGACTTCGAT	120
Db	61	TCAGAAAGACGTGAGAGATTTGTGACGCTTTTACGCGAACAGAAATATATACGACTTCGAT	120
QY	121	AAATCTTTTACGAGTCTCTGATCTTGACGAATGCAAAAAAACATGTTTCAAGACGGAG	180
Db	121	AAATCTTTTACGAGTCTCTGATCTTGACGAATGCAAAAAAACATGTTTCAAGACGGAG	180
QY	181	TACTGCTACATGTTTGTGTAAGACACGCGTCAACAAGAAATGTTACTCAATGTGCTTGAT	240
Db	181	TACTGCTACATGTTTGTGTAAGACACGCGTCAACAAGAAATGTTACTCAATGTGCTTGAT	240
QY	241	GGTGAAGAGTTGACCAAGAAATAATTTGTTGTCCACGAAATCTCACAGGAAATTTATG	300
Db	241	GGTGAAGAGTTGACCAAGAAATAATTTGTTGTGTCCACGAAATCTCACAGGAAATTTATG	300
QY	301	ACAGACTCGGAGGGTAAAGATGCAAGTAATCGGACAGGTACAGGTGACAGTCAGATGAA	360
Db	301	ACAGACTCGGAGGGTAAAGATGCAAGTAATCGGACAGGTACAGGTGACAGTCAGATGAA	360
QY	361	GTTGATGAGATTTAA	375
Db	361	GTTGATGAGATTTAA	375

RESULT 2	189967 bp	DNA	linear	HTG 17-JUL-2002
AL833802				
LOCUS				
DEFINITION	Mus musculus chromosome 4 clone RP23-17A4, *** SEQUENCING IN			
ACCESSION	AL833802			
VERSION	AL833802.3	GI:21911787		
KEYWORDS	HTG; HTGS; PHASE1.			
SOURCE	Mus musculus.			
ORGANISM	Mus musculus.			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 189967)			
AUTHORS	Plumb, B.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,			
	Cambridgeshire, CB10 1SA, UK. E-mail enquiries:			
	humuery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk			
	On Jul 19, 2002 this sequence version replaced gi:21758804.			
COMMENT	----- Genome Center			
	Center: Wellcome Trust Sanger Institute			

```

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquetry@sanger.ac.uk
-----
Project Information
Center project name: BM17A4
-----
Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 166623 bases at least Q40
Consensus quality: 167480 bases at least Q30
Consensus quality: 168086 bases at least Q20
Insert size: 188767; sum-of-contigs
Insert size: 215919; 10.9% error; agarose-fp
Quality coverage: 5.28x in Q20 bases; sum-of-contigs Quality
coverage: 4.68x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1. 8454: contig of 8454 bp in length
* 8455 8554: gap of 100 bp
* 8555 19307: contig of 10753 bp in length
* 19308 19407: gap of 100 bp
* 19408 27097: contig of 7690 bp in length
* 27098 27197: gap of 100 bp
* 27198 40822: contig of 13625 bp in length
* 40823 40922: gap of 100 bp
* 40923 48761: contig of 7839 bp in length
* 48762 48861: gap of 100 bp
* 48862 94932: contig of 46071 bp in length
* 94933 95032: gap of 100 bp
* 95033 109692: contig of 14660 bp in length
* 109693 109792: gap of 100 bp
* 109793 122076: contig of 12284 bp in length
* 122077 122176: gap of 100 bp
* 122177 130769: contig of 8593 bp in length
* 130770 130869: gap of 100 bp
* 130870 146267: contig of 15398 bp in length
* 146268 146367: gap of 100 bp
* 146368 151984: contig of 5617 bp in length
* 151985 152084: gap of 100 bp
* 152085 184112: contig of 32028 bp in length
* 184113 184212: gap of 100 bp
* 184213 189967: contig of 5755 bp in length.
*
* Location/Qualifiers
*
* 1. 189967
* /organism="Mus musculus"
* /db_xref="taxon:10090"
* /chromosome="4"
* /clone="RP23-17A4"
* /clone_lib="RPC1-23"
*
* 1. 8454
* /note="assembly_fragment:01243
* fragment_chain:1
* clone_end:SP6
* vector_side:left"
* 8555..19307
* /note="assembly_fragment:00590
* fragment_chain:1"
* 19408..27097
* /note="assembly_fragment:00259
* fragment_chain:1"
* 27198..40822
* /note="assembly_fragment:01468
* fragment_chain:1"
* 40923..48761
* /note="assembly_fragment:00793
* fragment_chain:2"

```

```

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquetry@sanger.ac.uk
-----
Project Information
Center project name: BM17A4
-----
Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 166623 bases at least Q40
Consensus quality: 167480 bases at least Q30
Consensus quality: 168086 bases at least Q20
Insert size: 188767; sum-of-contigs
Insert size: 215919; 10.9% error; agarose-fp
Quality coverage: 5.28x in Q20 bases; sum-of-contigs Quality
coverage: 4.68x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1. 8454: contig of 8454 bp in length
* 8455 8554: gap of 100 bp
* 8555 19307: contig of 10753 bp in length
* 19308 19407: gap of 100 bp
* 19408 27097: contig of 7690 bp in length
* 27098 27197: gap of 100 bp
* 27198 40822: contig of 13625 bp in length
* 40823 40922: gap of 100 bp
* 40923 48761: contig of 7839 bp in length
* 48762 48861: gap of 100 bp
* 48862 94932: contig of 46071 bp in length
* 94933 95032: gap of 100 bp
* 95033 109692: contig of 14660 bp in length
* 109693 109792: gap of 100 bp
* 109793 122076: contig of 12284 bp in length
* 122077 122176: gap of 100 bp
* 122177 130769: contig of 8593 bp in length
* 130770 130869: gap of 100 bp
* 130870 146267: contig of 15398 bp in length
* 146268 146367: gap of 100 bp
* 146368 151984: contig of 5617 bp in length
* 151985 152084: gap of 100 bp
* 152085 184112: contig of 32028 bp in length
* 184113 184212: gap of 100 bp
* 184213 189967: contig of 5755 bp in length.
*
* Location/Qualifiers
*
* 1. 189967
* /organism="Mus musculus"
* /db_xref="taxon:10090"
* /chromosome="4"
* /clone="RP23-17A4"
* /clone_lib="RPC1-23"
*
* 1. 8454
* /note="assembly_fragment:01243
* fragment_chain:1
* clone_end:SP6
* vector_side:left"
* 8555..19307
* /note="assembly_fragment:00590
* fragment_chain:1"
* 19408..27097
* /note="assembly_fragment:00259
* fragment_chain:1"
* 27198..40822
* /note="assembly_fragment:01468
* fragment_chain:1"
* 40923..48761
* /note="assembly_fragment:00793
* fragment_chain:2"

```

```

misc-feature      48862..94932
                  /note="assembly-fragment:00456
                  fragment.chain:2"
misc-feature      95033..109692
                  /note="assembly-fragment:00530
                  fragment.chain:2"
misc-feature      109793..122076
                  /note="assembly-fragment:01290
                  fragment.chain:2"
misc-feature      122177..130769
                  /note="assembly-fragment:01504.0"
                  /note="assembly-fragment:01504.1"
misc-feature      130870..146267
                  /note="assembly-fragment:01504.1"
misc-feature      146368..151984
                  /note="assembly-fragment:00822
                  fragment.chain:3"
misc-feature      152085..184112
                  /note="assembly-fragment:00663
                  fragment.chain:3"
misc-feature      184213..189967
                  /note="assembly-fragment:00178
                  fragment.chain:3
                  clone_end:T7
                  vector_side:right"

```

BASE COUNT 59392 a 34302 c 33745 g 61328 t 1200 others

Query Match 13.0%; Score 48.6; DB 2; Length 189967;
 Best Local Similarity 48.4%; Pred. No. 0.066; Mismatches 144; Indels 0; Gaps 0;

```

QY 97 AACGAGAAATATACAGACTTCGATTAATCTTTAAGAACTCCTGATCTTGACGAATGC 156
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 88182 AACAAAGCGCCGAGATTTGAACAGTTCTTAACAATAAAGAACTTTGGGGAATC 88241
QY 157 AAAAAAATGTTTCAAGACGAGTACTGCTACATCGTTTGAAGACACGGTCAACAAG 216
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 88242 ACCATCCATGACTTCAAGCTATCTACAGACAAATGATGATAGAGAAAGAAAGAA 88301
QY 217 GAATGTTACTACATGCTGCTGATGAGAGTTAGACCAAGAAATTTGTTGTCGAC 276
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 88302 GAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 88361
QY 277 GAAACCTTACGGAATAATTTATTTGACAGACTCGAGGGTAAAGATGCAAGTATCGGCA 336
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 88362 GAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 88421
QY 337 GGTACAGGTGACGATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 88422 GATTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 88460

```

RESULT 3
 AL807815/c 206914 bp DNA linear ROD 22-AUG-2002
 LOCUS Mouse DNA sequence from clone RP23-207H16 on chromosome 4, complete
 DEFINITION sequence.
 ACCESSION AL807815
 VERSION AL807815.3 GI:22474427
 KEYWORDS HTG.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORs Leonamornlert,D.
 TITLE Direct Submission
 JOURNAL Submitted (22-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Aug 23, 2002 this sequence version replaced gi:21690116.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute

COMMENT

Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-207H16 is from the RP23 Mouse PAC Library

constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6.

FEATURES
 Location/Qualifiers
 1..206914
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="4"
 /clone="RP23-207H16"
 /clone_lib="RP23-207H16"

BASE COUNT 64113 a 37887 c 37484 g 67430 t

Query Match 13.0%; Score 48.6; DB 10; Length 206914;
 Best Local Similarity 48.4%; Pred. No. 0.067; Mismatches 144; Indels 0; Gaps 0;

```

QY 97 AACGAGAAATATACAGACTTCGATTAATCTTTAAGAACTCCTGATCTTGACGAATGC 156
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 205163 AACAAAGCGCCGAGATTTGAACAGTTCTTAACAATAAAGAACTTTGGGGAATC 205104
QY 157 AAAAAAATGTTTCAAGACGAGTACTGCTACATCGTTTGAAGACACGGTCAACAAG 216
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 205103 ACCATCCATGACTTCAAGCTATCTACAGACAAATGATGATAGAGAAAGAAAGAA 205044
QY 217 GAATGTTACTACATGCTGCTGATGAGAGTTAGACCAAGAAATTTGTTGTCGAC 276
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 205043 GAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 204984
QY 277 GAAACCTTACGGAATAATTTATTTGACAGACTCGAGGGTAAAGATGCAAGTATCGGCA 336
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 204983 GAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 204924
QY 337 GGTACAGGTGACGATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 204923 GATTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 204885

```

RESULT 4
 AC006611 40090 bp DNA linear INV 19-APR-2002
 LOCUS Caenorhabditis elegans cosmid G30F8, complete sequence.
 DEFINITION AC006611
 ACCESSION AC006611
 VERSION AC006611.1 GI:4263327
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 I (bases 1 to 40090)

REFERENCE

AUTHORS
Waterston, R.
Genome sequence of the nematode *C. elegans*: a platform for
investigating biology. The *C. elegans* Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)

JOURNAL
MEDLINE
98069613
PUBMED
9851916

REFERENCE
2 (bases 1 to 40090)
Latreil, P., Twyman, B. and Wilson, R.
The sequence of *C. elegans* cosmid C30F8
Unpublished (2001)

JOURNAL
REFERENCE
3 (bases 1 to 40090)
Waterston, R.H.
Direct Submission
Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
4 (bases 1 to 40090)
Waterston, R.H.
Direct Submission
Submitted (01-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
5 (bases 1 to 40090)
Waterston, R.
Direct Submission
Submitted (14-AUG-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

REFERENCE
6 (bases 1 to 40090)
Waterston, R.
Direct Submission
Submitted (18-OCT-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

REFERENCE
7 (bases 1 to 40090)
Waterston, R.
Direct Submission
Submitted (12-FEB-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

REFERENCE
8 (bases 1 to 40090)
Waterston, R.
Direct Submission
Submitted (19-APR-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 0RO, England
email: rw@nematode.wustl.edu and jesse@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one m13 subclone.

For a graphical representation of this cosmid sequence and its
analysis see:
[\[www.wormbase.org/db/seq/sequence?name=C30F8;class=Sequence\]](http://www.wormbase.org/db/seq/sequence?name=C30F8;class=Sequence)

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C46H11, 8300 bp overlap; the 3' cosmid is Y110A7A,

FEATURES

source

200 bp overlap. Actual start of this cosmid is at base position 197
of C30F8; actual end is at 40090 of C30F8.

NOTES:
Coding sequences below are the result of integration and manual
review of the following data: computer analysts using the program
GeneFinder (P. Green and L. Hillier, personal communication), the
large scale EST projects of Yujin Kohara
(http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and the *C. elegans* ORFome cloning project (<http://wormfdb.dfci.harvard.edu/>),
similarity to other proteins from Blast analyses
(<http://blast.wustl.edu/>), sequence conservation with *C. briggsae*
using Jim Kent's WABA alignment program (Genome Research
10:1115-1125, 2000), individual *C. elegans* Genbank submissions,
and personal communications with *C. elegans* researchers. tRNAs
are predicted using the program tRNAscan-SE (Lowe, T.M. and
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

Location/Qualifiers

1..40090

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="I"

/clone="C30F8"

/complement(19565..21512)

/gene="C30F8.2"

/note="for a graphical representation of this gene see:
{www.wormbase.org/db/seq/sequence?name=C30F8.2;class=Seque
nce}"

/complement(join(19565..19654,20030..20101,20153..20407,
20460..20617,20879..21057,21102..21273,21392..21512))

/gene="C30F8.2"

/note="coded for by the following *C. elegans* cDNAs:

YK71092.3, YK449d2.3, YK705g11.3, YK745d7.3, YK169p6.3,
YK809d10.3, YK807d40.3, YK702g3.3, YK105h9.5,
YK169b6.5, YK306e11.3, YK306e11.5, YK431f1.3, YK431f1.5,
YK449d2.5, YK4J9e6.5, YK859e05.5, YK71092.5, YK702g3.5,
YK888h07.5, YK745d7.5, YK765e04.5, YK821e09.5, YK809d10.5,
YK733c05.5, YK705g11.5, YK754e11.5, YK807a04.5,
YK754e11.3, YK888h07.3, YK733c05.3, YK821a09.3"

/codon_start=1

/product="Hypothetical protein C30F8.2"

/protein_id="AAK85455.2"

/db_xref="GI:16259229"

/translation="MGEPLFNDHGLTALRGKGLLAADVANIYVQCELTEDKL
HISTDYGNFLANEPGATVYVIDEKLKLVETFTLRNNALBPLAFIDYIYSY
IDNILLITGLTHORPISLILNKCHLSEFDEMEAIHASTPAELYNVLDYPLANY
FVDCINEDDIDMEYEVIRNTLYKAYIDEDYKFCGAGGKTAEVDCILAEADRSI
ITVNSFDLSEKDRÖKLYPRGKLFPGDGLTGLSRADYDQVQVCEYFIDYRPE
GSGNGPGEKTLDEKFEHEVKLVHSHYLDHGFVAFILKLEDEMNIIMIAECIS
QHRRTKIDNYIPIIM"

/complement(21528..21529)

/note="SL1 trans-splice site: see YK821a09.3"

/complement(21534..21535)

/note="SL1 trans-splice site: see YK809d10.5"

25607..32891

/gene="C30F8.4"

/note="for a graphical representation of this gene see:
{www.wormbase.org/db/seq/sequence?name=C30F8.4;class=Sequ
ence}"

/join(25607..25684,27245..27344,27393..27492,27665..27859,
28346..28436,28489..28707,28818..28897,29946..30020,
30070..30315,32038..32325,32760..32891)

/gene="C30F8.4"

/note="similar to protein kinases (PFAM: pkinase, score;
183.25); coded for by the following *C. elegans* cDNAs:
YK739d9.5, YK284c11.3, YK284c11.5, YK465a7.5, YK739d9.3,
YK649h5.5"

/codon_start=1

/product="Hypothetical protein C30F8.4a"

/protein_id="AAK85457.1"

/db_xref="GI:15150650"

/db_xref="GI:15150650"

/db_xref="GI:15150650"

/db_xref="GI:15150650"

/db_xref="GI:15150650"

/db_xref="GI:15150650"

/db_xref="GI:15150650"

/db_xref="GI:15150650"

/db_xref="GI:15150650"

/db_xref="GI:15150650"

CDS

/translation="MEGLARVFLIGSKAVARYDEOTTIRVHVAVRGICISOVAA
HEALRTVSPOTAGSDSLMPLMRLITOLPHIYARHLPIVCEBRIEMRMP
OSVYLOATSSAFVYHIEVYDEFEHVAWSSVAVLALVKCRDAEHOHNG
ADHLEDLIDELACQSLIPVNLHPGKSHLKTFTAYIKRSATSPESILATL
LLEVKFDELVEFASLGAWTKPELVGHTLSYLRNCRDSSRLLEKTAETI
RKMGSEKTLMOILNGAPVILITSTELQSALHLDGOMLYNODSVKFKG
IERCELTMEATIRPTPNNDISNIRELLITLLELIGOGFVYKAVYDLKRD
ERIAVAVKCTAEAPADTOLIOESSIMRPHSNITOLIGCVOPMVLIELAPK
GELREYLOEKDMLPLRITLFCSDISLVYHSTRFVRDIAANILVCSPOCYL
ADRELSRALDYAVYTRSRKLPITKMLAPSVYKOPSMASDVMTGVCWELFSLG
KPMAGVNSDVIMHIEGSRPCPEKCPALYFISKMALEPHKRPVDOYALIE
DVROQIIONIPPOIIVGKPMIAGVIVAMSSLPGLITYRTMEDOKROEDAKMLE
QOEDEDEDODIDIPSTSHSVENIRTSNGYHHTPTSTRSLRFEDKTRGLRSVD
GVCDAVTKLONSNLHNDPFLHSYKVEVSOLEMLIVASGMRDVTITTOEDDM
TKTLIANDKOMSRVNGKLVNGHQAATNTLRDVRVIRGELAVNCTTIOLOLTPPL
ENESSILSNC"
join(27422..27492,27665..27859,28346..28436,28489..28707,
28818..29897,29946..30020,30070..30315,32038..32325,
32760..32891)
/gene="C30F8.4"
/note="coded for by the following C. elegans CDNs:
YK1020h08.5"
/codon_start=1
/product="Hypothetical protein C30F8.4b"
/protein_id="AA01554.1"
/db_xref="GI:20198781"
/translation="MLRITOLPHIYARHLPIVCEBRIEMRMPVSVYLOATSD
SAFVYLOEYVDEFEHVAWSSVAVLALVKCRDAEHOHNGADHLEDLIE
ACIOSLIPVNLHPGKSHLKTFTAYIKRSATSPESILATLLEVKFDEL
FKASLGAWTKPELVGHTLSYLRNCRDSSRLLEKTAETIRMGSEKTEL
MOLNLSGAAPVILITSTELQSALHLDGOMLYNODSVKFKGIERCELTME
ATIRPTPNNDISNIRELLITLLELIGOGFVYKAVYDLKRDIAVAVKCT
TAEAPADTOLIOESSIMRPHSNITOLIGCVOPMVLIELAPKGELETRYLO
DKMLPLRITLFCSDISLVYHSTRFVRDIAANILVCSPOCYLADRELSRA
LDYAVYTRSRKLPITKMLAPSVYKOPSMASDVMTGVCWELFSLGKPMAG
VNSDVIMHIEGSRPCPEKCPALYFISKMALEPHKRPVDOYALIEQVROQ
IIONIPPOIIVGKPMIAGVIVAMSSLPGLITYRTMEDOKROEDAKMLEQ
OEDEDEDODIDIPSTSHSVENIRTSNGYHHTPTSTRSLRFEDKTRGLRS
VDGVCDAVTKLONSNLHNDPFLHSYKVEVSOLEMLIVASGMRDVTITTO
EDDMTKTLIANDKOMSRVNGKLVNGHQAATNTLRDVRVIRGELAVNCTTIO
LOLTPPLENESSILSNC"

gene
35856..39749
/gene="C30F8.3"
/note="for a graphical representation of this gene see:
12.4%; Score 46.4; DB 3; Length 40090;
Best Local Similarity 52.6%; Pred. No. 0.23;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
Query Match
7 TATTCTGATTCCTCTTGGCTGCGAAGCTGCTGATCTCAACTCTTCAGAA 66
DB 8523 TTTCTCTTTTTCATATACATTTCTCAAAATTTTGAGCTTCTGACTCTCAAAAA 8582
OY 67 GAACGTGAAGATTGTTGAGCTTTACGACAGAAATAATACAGCTTCGATAATCT 126
DB 8583 TTAATGCTAGACTTTTAATCATTTTAAAGGTCAAAATAATTAATTCGAAATTTT 8642
OY 127 TTTAAGAATCCTCTGATCTTACGAGATGCAAAAAACATGTTCAAGAGGAGTACTGC 186
DB 8643 TTTGAGAAATTTCTTATCTGAAAAATCCTTACCAAAATTAATTCAGATGTTTTTTC 8702
OY 187 TACATGCTTTT 198
DB 8703 AAAAAATTTTTT 8714

RESULT 5
LOCUS AC006845 299670 bp DNA linear HTG 24-FEB-1999
DEFINITION Caenorhabditis elegans clone y119c1x, *** SEQUENCING IN PROGRESS
ACCESSION AC006845
VERSION AC006845.1 GI:4263503

KEYWORDS HTG, HTGS, PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 299670)
Waterson, R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 299670)
Waterson, R.H.
REFERENCE 2 (bases 1 to 299670)
Waterson, R.H.
JOURNAL Submitted (24-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2037: contig of 2037 bp in length
* 2038 2052: gap of unknown length
* 2053 14442: contig of 12390 bp in length
* 14443 14457: gap of unknown length
* 14458 101253: contig of 86796 bp in length
* 101254 101268: gap of unknown length
* 101269 264135: contig of 162867 bp in length
* 264136 264150: gap of unknown length
* 264151 299670: contig of 35520 bp in length.

FEATURES

source

1..299670
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y119C1X"

BASE COUNT 96582 a 52713 c 52700 g 97615 t 60 others
ORIGIN

Query Match
Best Local Similarity 52.6%; Pred. No. 0.25;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
OY 7 TATTCTGATTCCTCTTGGCTGCGAAGCTGCTGATCTCAACTCTTCAGAA 66
DB 211095 TTTCTCTTTTTCATATACATTTCTCAAAATTTTGAGCTTCTGACTCTCAAAAA 211154
OY 67 GAACGTGAAGATTGTTGAGCTTTACGACAGAAATAATACAGCTTCGATAATCT 126
DB 211155 TTAATGCTAGACTTTTAATCATTTTAAAGGTCAAAATAATTAATTCGAAATTTT 211214
OY 127 TTTAAGAATCCTCTGATCTTACGAGATGCAAAAAACATGTTCAAGAGGAGTACTGC 186
DB 211215 TTTGAGAAATTTCTTATCTGAAAAATCCTTACCAAAATTAATTCAGATGTTTTTTC 211274
OY 187 TACATGCTTTT 198
DB 211275 AAAAAATTTTTT 211286

RESULT 6
LOCUS AC115614/c 4458 bp DNA linear HTG 21-MAR-2002
DEFINITION Dictyostellium discoideum chromosome 2 map 5179259-5183715 strain
AX4, *** SEQUENCING IN PROGRESS ***; in ordered pieces.
ACCESSION AC115614
VERSION AC115614.1 GI:19570143
KEYWORDS HTG; HTGS, PHASE2.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
REFERENCE 1 (bases 1 to 4458)
Eukaryota; Mycetozoa; Dictyostellidae; Dictyostellium.

LOCUS	DEFINITION	AP004782	119191 bp	DNA	linear	PRI 11-JUL-2002
AP004782	Homo sapiens genomic DNA, chromosome 11 clone: CMB9-103B16, complete sequence.	AP004782	GI:21728157			
ACCESSION	VERSION	AP004782.3	GI:21728157			
KEYWORDS	SOURCE	HTG.				
ORGANISM		Homo sapiens DNA, clone: CMB9-103B16.				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS		Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.				
TITLE		Homo sapiens genomic DNA				
JOURNAL		Published Only in Database (2002)				
REFERENCE		2 (bases 1 to 119191)				
AUTHORS		Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.				
TITLE		Direct Submission				
JOURNAL		Submitted (25-FEB-2002) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Shinto-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: hattori@psc.riken.go.jp, URL: http://ngp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)				
COMMENT		On Jul 10, 2002 this sequence version replaced gi:21328197.				
FEATURES		Location/Qualifiers				
SOURCE		1. .119191				
		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
		/chromosome="11"				
		/map="11q"				
		/clone="CMB9-103B16"				
BASE COUNT		34204 a 26845 c 25557 g 32585 t				
ORIGIN						
Query Match		12.2%; Score 45.8; DB 9; Length 119191;				
Best Local Similarity		53.7%; Pred. No. 0.34; Matched 82; Indels 0; Gaps 0;				
Matches		95; Conservative 0; Mismatches 82; Indels 0; Gaps 0;				
QY	199	GAAGACACGCGTCACCAAGAAAGTGTACTACATGTCGTGATGGTGAAGAGTTAGACCA	258			
Db	102658	GAAAGCAAGAGACGACGACGAGAAAGACGACGACGACGACGACGACGACGACGACGACG	102717			
QY	259	GAAAAATTTGTTGTCGACGACGAAACTTCACGCGAAATTTTTCACAGACTCGGAGGCTAA	318			
Db	102718	GAAAGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	102777			
QY	319	GATGACGAGTATGCGGCGCAGGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTATTA	375			
Db	102778	GAAAGCAAGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	102834			
RESULT 10						
LOCUS		AF312994	228283 bp	DNA	linear	ROD 17-JAN-2002
DEFINITION		Mus musculus chromosome 1 clone MML, complete sequence.				
ACCESSION		AF312994				
VERSION		AF312994.1	GI:18182288			
KEYWORDS		HTG.				
SOURCE		Mus musculus.				
ORGANISM		Mus musculus				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.				
AUTHORS		Rump, A., Hess, U., Galgoczy, P., Wirth, T. and Rosenthal, A.				
TITLE		Genomic sequence from mouse chr. 1				
JOURNAL		Unpublished				
REFERENCE		2 (bases 1 to 228283)				
AUTHORS		Rump, A., Hess, U., Galgoczy, P., Wirth, T. and Rosenthal, A.				
TITLE		Direct Submission				
JOURNAL		Submitted (12-OCT-2000) Genome Analysis, Institute of Molecular				

FEATURES	Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany				
SOURCE	location/Qualifiers				
	1..228283				
	/organism="Mus musculus"				
	/db_xref="taxon:10090"				
	/chromosome="1"				
	/clone="NML"				
BASE COUNT	63688 a	49622 c	49154 g	65819 t	
ORIGIN					
Query Match	12.2%	Score 45.8;	DB 10;	Length 228283;	
Best Local Similarity	49.8%;	Pred. No. 0.35;			
Matches 116; Conservative	0;	Mismatches 117;	Indels 0;	Gaps 0;	
QY 143	ATCTTGACGATGCAAAAAACATGTTTCAAGACGAGTACTGCTACATCGTTTGTGAAG	202			
Db 148716	ATTTTGAAGAAAAAAATCATATTATTACAGATTCCAGACTGTGTATATGAAACAAG	148775			
QY 203	ACACGGTCACACAGGAATGTTACTACATGCTGCTGTGATGGTGAAGAGTTAGACCAAGAA	262			
Db 148776	AAGAAG	148835			
QY 263	AATTTGTGTCGAGAAAACCTTCACGGAATAATTTTTCACAGACTGCGAGGCTAAAGATG	322			
Db 148836	AAGAAG	148895			
QY 323	CAGTAATGCGCGCAGGTACAGGTGACGATCGATGATGAAGTTGATGAAGATTAA	375			
Db 148896	AAGAAG	148948			
RESULT 11					
AF216973	20796 bp	DNA	linear	INV 17-FEB-2000	
LOCUS	Drosophila melanogaster 1(2)01289 gene, complete cds, alternatively				
DEFINITION	spliced products.				
ACCESSION	AF216973				
VERSION	AF216973.1	GI:6984064			
KEYWORDS					
SOURCE	Drosophila melanogaster.				
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1. (bases 1 to 20796)				
AUTHORS	Manning, G., Micklem, D. R. and Krasnow, M. A.				
JOURNAL	Sequence of 1(2)01289 genomic locus				
REFERENCE	Unpublished				
TITLE	2 (bases 1 to 20796)				
JOURNAL	Manning, G., Micklem, D. R. and Krasnow, M. A.				
REFERENCE	Submitted (16-DEC-1999) Department of Biochemistry, Stanford				
AUTHORS	University, Beckman Center, Stanford, CA 94305, USA				
TITLE	Location/Qualifiers				
JOURNAL	1..20796				
FEATURES	/organism="Drosophila melanogaster"				
Source	/db_xref="taxon:7237"				
gene	<865..20489				
mRNA	/gene="1(2)01289"				
	join(<865..1470,3154..3310,3382..3548,3980..4015,6598..6780,6911..7054,7117..7250,7315..7495,7567..7881,8252..8566,8627..9271,9325..9507,9575..9712,9776..9958,15122..15253,15483..15806,15871..16059,16118..16255,17488..17670,17723..17854,18603..18759,18817..18983,19047..19203,19257..19423,19484..19666,19733..20489)/gene="1(2)01289"				
	/product="1(2)01289 long form"				
	join(<1153..1470,3154..3310,3382..3548,8252..8566,8627..9271,9325..9507,9575..9712,9776..9958,10019..12437..12919,13190..14706)/gene="1(2)01289"				
	/product="1(2)01289 H9 form"				
	join(1257..1470,3154..3310,3382..3548,3980..4015,				
CDS					

[illegible]

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 13:13:01 ; Search time 218 Seconds
(Without alignments)
3873.850 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 375

Sequence: 1 atgaagatattcttgattc.....atgaagtgtatgaattaa 375

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	375	100.0	375	21	AAA28180
2	45.8	12.2	49999	20	AAZ23891
3	45.8	12.2	49999	20	AAZ23896
4	45.2	12.1	15935	23	ABU06114
5	44.4	11.8	2438	24	ABO54916
6	41	10.9	372	23	AA575452
7	40.2	10.7	3211	18	AA89346
8	40.2	10.7	3901	18	AA89345
9	40	10.7	2365589	24	ABA90521

C	10	39.6	10.6	766	22	ABA49279	Human breast cell
C	11	39.6	10.6	766	22	ABA67193	Human foetal liver
C	12	39.6	10.6	766	22	ABA34288	Probe #12754 for g
C	13	39.6	10.6	766	22	AA122114	Probe #12047 for g
C	14	39.6	10.6	766	22	AA147409	Probe #16095 used
C	15	39.6	10.6	766	22	AA107812	Probe #7803 used t
C	16	39.6	10.6	1944	22	ABA44123	Human breast cell
C	17	39.6	10.6	1944	22	ABA54575	Human foetal liver
C	18	39.6	10.6	1944	22	ABA24359	Probe #2825 for ge
C	19	39.6	10.6	1944	22	AA112874	Probe #2807 for ge
C	20	39.6	10.6	1944	22	AA134231	Probe #2917 used t
C	21	39.6	10.6	1944	22	AA102791	Kaposi's sarcoma-a
C	22	39.6	10.6	3489	21	AAA30290	Nucleotide sequenc
C	23	39.6	10.6	3489	22	AA893901	Kaposi's sarcoma-a
C	24	39.6	10.6	3489	24	ABA93487	KSHV LTR DNA (nucl
C	25	39.6	10.6	32207	20	AAV73805	KSHV Long unique c
C	26	39.6	10.6	137507	19	AAV19941	DNA encoding novel
C	27	39.4	10.5	354	23	AA590677	DNA encoding novel
C	28	39.4	10.5	654	23	AA569553	DNA encoding novel
C	29	39.4	10.5	654	23	AA571153	DNA encoding novel
C	30	39.4	10.5	654	23	AA575467	DNA encoding novel
C	31	39.2	10.5	234	23	AA575455	DNA encoding novel
C	32	39	10.4	400	23	AA575460	DNA encoding novel
C	33	39	10.4	963	23	AA568580	DNA encoding novel
C	34	38.6	10.3	1448	21	AAAC47605	Arabidopsis thalia
C	35	38.4	10.2	1072	23	AA590738	DNA encoding novel
C	36	37.8	10.1	2944528	24	ABA93041	Listeria monocytog
C	37	37.6	10.0	1622	23	AA564269	DNA encoding novel
C	38	37.4	10.0	2658	21	AAA70213	Plasmodium falcipa
C	39	37.4	10.0	3543	22	AAH62819	Shrimp white spot
C	40	37.4	10.0	305107	22	AAH62689	Shrimp white spot
C	41	37.2	9.9	315	22	ABA51452	Human breast cell
C	42	37.2	9.9	315	22	ABA69479	Human foetal liver
C	43	37.2	9.9	315	22	ABA36403	Probe #14869 for g
C	44	37.2	9.9	315	22	AAK17735	Human brain expres
C	45	37.2	9.9	315	22	AAK43555	Human bone marrow

ALIGNMENTS

RESULT 1	AAA28180	standard; cDNA; 375 BP.
ID	AAA28180	
AC	AAA28180	
XX		
DT	29-JAN-2001	(first entry)
XX		
DE	Platelet binding inhibitor protein Saratin encoding cDNA sequence.	
KW	Saratin; medicinal leech; thromboembolic disease; intraocular lens;	
KW	collagen-dependent platelet adhesion inhibition;	
KW	posterior capsule opacification; ss.	
XX		
OS	Hirudo medicinalis.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	64..375
FT		/tag= a
FT		/product= "Saratin"
FT		/note= "Platelet adhesion inhibitor protein"
PN	WO200056885-A1.	
XX		
PD	28-SEP-2000.	
XX		
PF	10-MAR-2000; 2000WO-EP02117.	
XX		
PR	18-MAR-1999; 99EP-0105530.	
XX		
PR	12-MAY-1999; 99EP-0109503.	
XX		
PA	(MERCK) MERCK PATENT GMBH.	

CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 372 BP; 194 A; 27 C; 134 G; 17 T; 0 other;

Query Match 10.9%; Score 41; DB 23; Length 372;
 Best Local Similarity 48.9%; Pred. No. 0.043;
 Matches 110; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

OY 151 GAATCAAAAAACATGTTCAACGAGGAGTCTACATCGTTTGAAGACACGGTC 210
 DB 31 GAATCCCAATAAACGGGAGTTGAGAGCTCAGTCAACATAGCAACCCCATCTGAAA 90
 OY 211 AACAAAGAAATGTTACTACATGTCGTTGATGTAAGAGTTAGACCAAGAAAAATTTGTT 270
 DB 91 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 150
 OY 271 GTGACGAAAACTCAGCGAAATTTATTTGACAGCTGCGAGGGTAAAGATGACGTAAT 330
 DB 151 GAGAGAGAAAGAAAG 210
 OY 331 GCGGAGGTACAGGTGACGAGTCAAGTCAAGTTGATGAAGATTAA 375
 DB 211 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 255

RESULT 7
 AAT89346
 ID AAT89346 standard; cDNA; 3211 BP.

XX AC AAT89346;
 XX DT 11-MAR-1998 (first entry)
 XX DE Human p160 cDNA 160.2.

XX KW p160; p62; cytoplasmic; T cell; B cell; development; activation;
 KW modulation; cellular response; cell proliferation; autoimmune disease;
 XX p56-lck; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX FT CDS 439..3156
 XX FT /*tag= a
 XX FT /product= p160
 XX FT /note= "160.2"

XX PN W09722255-A1.

XX PD 26-JUN-1997.

XX PF 11-DEC-1996; 96WO-US19944.

XX PR 19-DEC-1995; 95US-0574959.

XX PA (DAND) DANA FARBER CANCER INST INC.

XX PI Young I, Shin J, Strominger JL, Vadamudi RK;

XX DR WPI: 1997-341351/31.

XX P-PSDB; AAW31186.

XX CDNA encoding p62 and p160 and corresponding proteins - used in the
 PT treatment of autoimmune disease and for T and B cell proliferation,
 PT e.g. for treatment of tumours

XX Claim 82; Fig 10; 175pp; English.

CC This cDNA sequence encodes a novel p160 (160.2) which is capable
 CC of activating transcription of a variety of genes upon activation of p62
 CC and is capable of binding to the p62/p56lck complex to modulate lck
 CC function in a manner similar to p62. The genes transcribed in response to
 CC p160 activation likely include those of which are involved in T or B cell
 CC development/differentiation, T or B cell activation or production of T or
 CC B cell specific factors e.g. lymphokines or antibodies. This p160
 CC polypeptide is also a substrate for serine/threonine kinase activity.
 CC p160 polypeptides can modulate degradation of cellular proteins e.g. cell
 CC cycle regulatory proteins stimulating expression of cell cycle dependent
 CC kinase inhibitors and arresting cell cycle progression at specific
 CC boundaries to thereby modulate cell proliferation. As p160 boosts B cell
 CC response it may be used to treat disorders where this is beneficial, e.g.
 CC infections by pathogenic microorganisms. p160 can be used to expand T
 CC cell populations for treating infectious diseases or cancer and p160
 CC inhibitors could reduce B or T cell responses and may be used to treat a
 CC variety of autoimmune diseases, e.g. diabetes mellitus, arthritis,
 CC multiple sclerosis allergic reactions, Crohn's diseases etc.

XX Sequence 3211 BP; 649 A; 990 C; 948 G; 624 T; 0 other;

Query Match 10.7%; Score 40.2; DB 18; Length 3211;
 Best Local Similarity 52.0%; Pred. No. 0.17; 83; Indels 0; Gaps 0;
 Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

OY 199 GAGACACGCTCAACAAGAAATGTTACTCAATGTCGTTGATGTAAGATTAGACAA 258
 DB 2458 GAGCAAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2517
 OY 259 GAAAAATTTGTTGTCGACGAAAACTTCACGAAAAATTTATTTGACAGACTGCGAGGTA 318
 DB 2518 GAATATTTTGAAG 2577
 OY 319 GATGCAAGTATCGGACAGGTACAGGTGACGATCAGATGAATGATTAAGA 371
 DB 2578 GAAAGTGAGTTAG 2630

RESULT 8
 AAT89345
 ID AAT89345 standard; cDNA; 3901 BP.

XX AC AAT89345;

XX DT 11-MAR-1998 (first entry)

XX DE Human p160 cDNA 160.1.

XX KW p160; p62; cytoplasmic; T cell; B cell; development; activation;
 KW modulation; cellular response; cell proliferation; autoimmune disease;
 XX p56-lck; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX FT CDS 439..3846
 XX FT /*tag= a
 XX FT /product= p160
 XX FT /note= "160.1"

XX PN W09722255-A1.

XX PD 26-JUN-1997.

XX PF 11-DEC-1996; 96WO-US19944.

XX PR 19-DEC-1995; 95US-0574959.

XX PA (DAND) DANA FARBER CANCER INST INC.

XX PI Young I, Shin J, Strominger JL, Vadamudi RK;

XX DR WPI: 1997-341351/31.

XX WPI: 2001-496933/54.
 DR New spatially-addressable set of single exon nucleic acid probes,
 XX useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes
 PS
 PS Claim 4; SEQ ID NO 7974; 327bp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other:
 Query Match 10.6%; Score 39.6; DB 22; Length 766;
 Best Local Similarity 51.1%; Pred. No. 0.14;
 Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 187 TACATCGTTTGAAGACAGCGTCAACAGATGTTACTACAAATGCTTGATGTGA 246
 DB 687 TCCACACTTGAGACACACCATGAGATGATGATGATGATGATGATGATGATG 628
 QY 247 GAGTTAGCCAAAGAAATTTGTTGTCGACGAAACTTCACGAAATTTTGGACAGAC 306
 DB 627 GATGAGAGAGATGATATATATGATGACGATGATGATGATGATGATGATGATG 568
 QY 307 TGCAGAGGTAAGATGACGATTAATGCGCAGGTACAGGTGACAGTCAATGATGAT 366
 DB 567 GACGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
 QY 367 GA 368
 DB 507 GA 506
 RESULT 11
 ABA67193/c
 ID ABA67193 standard; DNA; 766 BP.
 AC
 AC ABA67193;
 XX
 DT 01-FEB-2002 (first entry)
 DE Human foetal liver single exon nucleic acid probe #15498.
 DE Human foetal liver; gene expression; single exon nucleic acid probe; ss.
 KW
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 OS Homo sapiens.
 OS
 PN WO200157277-A2.
 PN
 PD 09-AUG-2001.
 PD
 PF 30-JAN-2001; 2001WO-US000669.
 PF
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR

PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234367.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 DR WPI: 2001-483447/52.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver
 PS
 PS Claim 4; SEQ ID NO 15498; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other:
 Query Match 10.6%; Score 39.6; DB 22; Length 766;
 Best Local Similarity 51.1%; Pred. No. 0.14;
 Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 187 TACATCGTTTGAAGACACGCTCAACAGATGTTACTACAAATGCTTGATGTGA 246
 DB 687 TCCACACTTGAGACACACCATGAGATGATGATGATGATGATGATGATGATG 628
 QY 247 GAGTTAGCCAAAGAAATTTGTTGTCGACGAAACTTCACGAAATTTTGGACAGAC 306
 DB 627 GATGAGAGAGATGATATATATGATGACGATGATGATGATGATGATGATGATG 568
 QY 307 TGCAGAGGTAAGATGACGATTAATGCGCAGGTACAGGTGACAGTCAATGATGAT 366
 DB 567 GACGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
 QY 367 GA 368
 DB 507 GA 506
 RESULT 12
 ABA34288/c
 ID ABA34288 standard; DNA; 766 BP.
 AC
 AC ABA34288;
 XX
 DT 23-JAN-2002 (first entry)
 DE Probe #12754 for gene expression analysis in human heart cell sample.
 DE Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 OS Homo sapiens.
 OS
 PN WO200157274-A2.
 PN
 PD 09-AUG-2001.
 PD
 PF 30-JAN-2001; 2001WO-US00066.
 PF
 PR 04-FEB-2000; 2000US-0180312.
 PR


```

XX 30-JAN-2001; 2001WO-US00663.
PF 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 16095; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other;
Query Match 10.6%; Score 39.6; DB 22; Length 766;
Best Local Similarity 51.1%; Pred. No. 0.14;
Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 187 TACATCGTTTGAAGACACGCTCAACAGCATGTTACTCATGTCGTTGATGGTAA 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 687 TCACACTTGAGGACACCATGAGATGATGATGATGATGATGATGATGATGATGAT 628
QY 247 GAGTTAGACCAAGAAAAATTTGTTGTCGACGAAACTTCACGGAATTTATTGACAGAC 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 627 GATGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 568
QY 307 TCGGAGGGTAAAGATGACGATTAATCGCGACGATACAGTGCAGATGAAATTGAT 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 567 GACGAGGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
QY 367 GA 368
    |||
DB 507 GA 506
RESULTS
AA107812/C
ID AA107812 standard; DNA; 766 BP.
XX
XX AA107812;
XX
XX 09-OCT-2001 (first entry)
XX
XX Probe #7803 used to measure gene expression in human breast sample.
DE
XX Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
OS
XX
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.

```

```

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
XX Claim 25; SEQ ID No 7803; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridises at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and
XX non-carcinoma tumours.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other;
Query Match 10.6%; Score 39.6; DB 22; Length 766;
Best Local Similarity 51.1%; Pred. No. 0.14;
Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 187 TACATCGTTTGAAGACACGCTCAACAGCATGTTACTCATGTCGTTGATGGTAA 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 687 TCACACTTGAGGACACCATGAGATGATGATGATGATGATGATGATGATGATGAT 628
QY 247 GAGTTAGACCAAGAAAAATTTGTTGTCGACGAAACTTCACGGAATTTATTGACAGAC 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 627 GATGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 568
QY 307 TCGGAGGGTAAAGATGACGATTAATCGCGACGATACAGTGCAGATGAAATTGAT 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 567 GACGAGGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
QY 367 GA 368
    |||
DB 507 GA 506

```

Search completed: May 5, 2003, 17:29:32
 Job time : 870 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 16:39:12 ; Search time 50 Seconds
(without alignments)
2300.076 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 375
Sequence: 1 atgaagattcttctgatttc.....atgaagtgtgaagatttaa 375

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_NA.*
2: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	42.2	11.3	7218	1 US-08-232-463-14	Sequence 14, Appl
2	40.2	10.7	3211	2 US-08-574-959A-8	Sequence 8, Appl1
3	40.2	10.7	3211	4 US-09-357-014-8	Sequence 8, Appl1
4	40.2	10.7	3901	4 US-08-574-959A-6	Sequence 6, Appl1
5	40.2	10.7	3901	4 US-09-357-014-6	Sequence 6, Appl1
6	39.6	10.6	3489	4 US-08-728-323A-1	Sequence 1, Appl1
7	39.6	10.6	3489	4 US-09-298-568-1	Sequence 1, Appl1
8	39.6	10.6	32207	2 US-08-770-379-20	Sequence 20, Appl
9	39.6	10.6	32207	4 US-08-757-669A-20	Sequence 20, Appl
C 10	39.6	10.6	32207	4 US-09-230-371A-20	Sequence 20, Appl
11	36	9.6	966	2 US-08-766-738-2	Sequence 2, Appl1
12	36	9.6	966	4 US-09-262-610-2	Sequence 2, Appl1
13	35	9.3	2694	2 US-08-867-941-2	Sequence 2, Appl1
14	35	9.3	2694	4 US-08-867-941-2	Sequence 2, Appl1
15	35	9.3	7650	2 US-08-867-941-1	Sequence 1, Appl1
16	35	9.3	7650	4 US-09-074-658-1	Sequence 1, Appl1
17	35	9.3	8920	2 US-08-446-855A-1	Sequence 1, Appl1
18	35	9.3	8920	4 US-08-446-855A-1	Sequence 1, Appl1
19	34.8	9.3	5361	4 US-09-150-741-1	Sequence 1, Appl1
20	34.8	9.3	6152	4 US-08-973-462-2	Sequence 1, Appl1
21	34.6	9.2	2518	4 US-08-973-462-1	Sequence 1, Appl1
22	34.6	9.2	5852	4 US-09-433-699-3	Sequence 3, Appl1
23	34	9.1	759	1 US-07-867-106-2	Sequence 2, Appl1
24	34	9.1	759	1 US-08-466-603-4	Sequence 4, Appl1
25	34	9.1	759	1 US-08-314-503A-4	Sequence 4, Appl1
26	34	9.1	759	2 US-08-468-066-4	Sequence 4, Appl1
27	34	9.1	759	3 US-08-466-717-4	Sequence 4, Appl1

28	34	9.1	759	5 PCT-US95-12414-4	Sequence 4, Appl1
29	34	9.1	980	1 US-08-466-603-3	Sequence 3, Appl1
30	34	9.1	980	1 US-08-314-503A-3	Sequence 3, Appl1
31	34	9.1	980	1 US-08-468-066-3	Sequence 3, Appl1
32	34	9.1	980	2 US-08-466-717-3	Sequence 3, Appl1
33	34	9.1	980	3 US-08-466-717-3	Sequence 3, Appl1
34	34	9.1	980	3 US-08-466-717-3	Sequence 3, Appl1
35	33.8	9.0	4518	4 US-08-961-527-121	Sequence 121, App
36	33.6	9.0	2295	1 US-08-375-300-3	Sequence 3, Appl1
37	33.6	9.0	2295	3 US-09-177-431-3	Sequence 3, Appl1
38	33.6	9.0	2295	5 PCT-US95-16930-3	Sequence 3, Appl1
39	33.6	9.0	4080	1 US-08-375-300-1	Sequence 1, Appl1
40	33.6	9.0	4080	3 US-09-177-431-1	Sequence 1, Appl1
41	33.6	9.0	4080	5 PCT-US95-16930-1	Sequence 1, Appl1
42	33.4	8.9	1137	4 US-09-134-001C-657	Sequence 657, App
43	33.4	8.9	1727	4 US-09-071-035-295	Sequence 295, App
44	33.4	8.9	1839	4 US-09-071-035-293	Sequence 293, App
45	33.2	8.9	4599	1 US-08-431-080-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-Fls
; US-08-232-463-14
Query Match 11.3%; Score 42.2; DB 1; Length 7218;

Best Local Similarity 4.6%; Pred. No. 0.0023;
Matches 14; Conservative 170; Mismatches 123; Indels 0; Gaps 0;

QY 65 AGACAGTGAAGATTGTTGACGTTTACGCAACAGAAATATACAGCTTCGTAAT 124
||||| ||||| :
Db 1456 AAGAGTAAAGAAATTTGTTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1397
QY 125 CTTTAAAGAGTCTGCTGTTGACGATGCAAAAAAATGTTCAAGAGAGTCT 184
:
Db 1396 RRR 1337
QY 185 GTCATACGTTTGAAGACAGGTCACAGAAATGTTACTACATGCTGTGATG 244
:
Db 1336 RRR 1277
QY 245 AAGAGTTAGCAACAAATTTGTTGTCAGCAAAATCTCAGCAAAATTTTTCAG 304
:
Db 1276 RRR 1217
QY 305 ACTGCGAGGTAAAGATGAGTAATGCGGACGATGACGATGATGAGTGA 364
:
Db 1216 RRR 1157
QY 365 ATGAGA 371
:
Db 1156 RRRRRR 1150

RESULT 2

US-08-574-959A-8
; Sequence 8, Application US/08574959A
; Patent No. 5962224

GENERAL INFORMATION:

APPLICANT: Jaekyoon Shin, Insil Jouny, Ratna K. Vadlamudi
and Jack L. Strominger
APPLICANT: and Jack L. Strominger
TITLE OF INVENTION: P62 POLYPEPTIDES, RELATED POLYPEPTIDES
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts

COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/574,959A
FILING DATE: 19-DEC-95

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 3211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS
LOCATION: 439..3157
US-08-574-959A-8

Query Match 10.7%; Score 40.2; DB 2; Length 3211;

Best Local Similarity 52.0%; Pred. No. 0.0066;
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 199 GAACACGCGTCAACAGAAATGTTACTACATCTCTGATGGTGAAGAGTTAGACCA 258
||||| ||||| :
Db 2458 GAAG 2517
QY 259 GAAATTTGTTGTCGAGAAATCTCAGCAAAATTTTGAACAGCTCGAGGTTAA 318
||||| ||||| :
Db 2518 GAATATTTTGAAG 2577
QY 319 GATCAGAGTAATGCGGACGATGACGAGTTCAGATGAAGTTGATGAGA 371
||||| ||||| :
Db 2578 GAAGTGAGTTAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2630

RESULT 3

US-09-357-014-8
; Sequence 8, Application US/09357014
; Patent No. 6291645

GENERAL INFORMATION:

APPLICANT: Jaekyoon Shin, Insil Jouny, Ratna K. Vadlamudi
and Jack L. Strominger
APPLICANT: and Jack L. Strominger
TITLE OF INVENTION: P62 POLYPEPTIDES, RELATED POLYPEPTIDES
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts

COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/357,014
FILING DATE: 19-Jul-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/574,959
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 3211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS
LOCATION: 439..3157
US-09-357-014-8

Query Match 10.7%; Score 40.2; DB 4; Length 3211;

Best Local Similarity 52.0%; Pred. No. 0.0066;
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 199 GAACACGCGTCAACAGAAATGTTACTACATCTCTGATGGTGAAGAGTTAGACCA 258
||||| ||||| :
Db 2458 GAAG 2517
QY 259 GAAATTTGTTGTCGAGAAATCTCAGCAAAATTTTGAACAGCTCGAGGTTAA 318

TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match
Best Local Similarity 47.2%; Score 39.6; DB 2; Length 3489;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAAATCTTTAAGAGCTCTGATCTTGACGAAATGCAAAAAACATGTTTCAAGAGC 177
DB 871 GAGAAGACATCCAGAGTAAAGAAATGTCGAAATATATAGCTGCGAGGATAT 930
QY 178 GAGTACGTCTACATCGTTTGAAGACGCGTCAACAGGAATGTTACTACATGTCGTT 237
DB 931 GGGACACAGAGATTAGCAAGAAAGTGCAGTGCACAGATGACATGATTAAGAT 990
QY 238 GATGTTGAAGATTAGACCAAGAAAAATTTGTTGCGACGAAACTTCACGGAATTTAT 297
DB 991 GATGAGGAGGAGCAGAGAGATGAGAGGAGAGAGATGACGAGAGATGACGAG 1050
QY 298 TTGACAGACTGCGAGGTAAGATGCGTAATGCGGACAGTACAGTGCAGTACAT 357
DB 1051 GAGGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAG 1110
QY 358 GAAGTTGATGAAGA 371
DB 1111 GAGGATGACGAGGA 1124

RESULT 7
US-09-298-568-1
Sequence 1, Application US/09298568
Patent No. 6322792
GENERAL INFORMATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kieft, Elliott D.
REGISTRATION NUMBER:
APPLICANT: Bailestas, Mary E.
APPLICANT: Kye, Kenneth M.
TITLE OF INVENTION: RHADINO VIRUS IANA ACTS IN TRANS ON A UNIT OF RHADINO
TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
FILE REFERENCE: 16412-10001R

CURRENT APPLICATION NUMBER: US/09/298,568
CURRENT FILING DATE: 1999-04-21
EARLIER APPLICATION NUMBER: US 60/109,422
EARLIER FILING DATE: 1998-11-19
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 3489
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1

Query Match
Best Local Similarity 47.2%; Score 39.6; DB 4; Length 3489;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAAATCTTTAAGAGCTCTGATCTTGACGAAATGCAAAAAACATGTTTCAAGAGC 177
DB 871 GAGAAGACATCCAGAGTAAAGAAATGTCGAAATATATAGCTGCGAGGATAT 930
QY 178 GAGTACGTCTACATCGTTTGAAGACGCGTCAACAGGAATGTTACTACATGTCGTT 237
DB 931 GGGACACAGAGATTAGCAAGAAAGTGCAGTGCACAGATGACATGATTAAGAT 990
QY 238 GATGTTGAAGATTAGACCAAGAAAAATTTGTTGCGACGAAACTTCACGGAATTTAT 297
DB 991 GATGAGGAGGAGCAGAGAGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAG 1050
QY 298 TTGACAGACTGCGAGGTAAGATGCGTAATGCGGACAGTACAGTGCAGTACAT 357
DB 1051 GAGGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAG 1110
QY 358 GAAGTTGATGAAGA 371
DB 1111 GAGGATGACGAGGA 1124

RESULT 8
US-08-770-379-20/c
Sequence 20, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 10.6%; Score 39.6; DB 2; Length 32207;
Best Local Similarity 47.2%; Pred. No. 0.03;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAATCTTTTAAAGACTCTCTGATCTTGACGAATGCAGAAAAACATGTTTCAAGCG 177
DB 21126 GAAGAAGCATCCAGAAATGAAAAAGAAATGTTCCGAAATTAATCAGCTGGCGAGATAT 21067
QY 178 GAGTACTGCTACATCGTTTGAAGACAGGTCAACAAGAAATGTACTACATGTGCTT 237
DB 21066 GGGGACAAACAGATTGACAAAGAAAGTGCAGGTGACAAAGATGACAAATAGAGAT 21007
QY 238 GATGTTGAAGAGTTAGACCAAGAAATTTGTTGCGACGAAACTTCACGAAATTTAT 297
DB 21006 GATGAGGAGAGACAGACAGACAGATGAGAGAGAGAGATGACGAGAGATGACGAG 20947
QY 298 TTGACAGACTGCGAGGTAAGATGATGATGCGGAGGTACAGGTGACGAGTACAGAT 357
DB 20946 GAGGATGACGAGAGAGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAG 20887
QY 358 GAAGTTGATGAGA 371
DB 20886 GAGGATGACGAGGA 20873

RESULT 9

US-08-757-669A-20/c
Sequence 20, Application US/08757669A
Patent No. 6183751
GENERAL INFORMATION:

APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757/669A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-F
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20

Query Match 10.6%; Score 39.6; DB 4; Length 32207;
Best Local Similarity 47.2%; Pred. No. 0.03;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAATCTTTTAAAGACTCTCTGATCTTGACGAATGCAGAAAAACATGTTTCAAGCG 177
DB 21126 GAAGAAGCATCCAGAAATGAAAAAGAAATGTTCCGAAATTAATCAGCTGGCGAGATAT 21067
QY 178 GAGTACTGCTACATCGTTTGAAGACAGGTCAACAAGAAATGTACTACATGTGCTT 237
DB 21066 GGGGACAAACAGATTGACAAAGAAAGTGCAGGTGACAAAGATGACAAATAGAGAT 21007
QY 238 GATGTTGAAGAGTTAGACCAAGAAATTTGTTGCGACGAAACTTCACGAAATTTAT 297
DB 21006 GATGAGGAGAGACAGACAGATGAGAGAGAGAGATGACGAGAGATGACGAGATGACGAG 20947
QY 298 TTGACAGACTGCGAGGTAAGATGATGATGCGGAGGTACAGGTGACGAGTACAGAT 357
DB 20946 GAGGATGACGAGAGAGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAGAG 20887
QY 358 GAAGTTGATGAGA 371
DB 20886 GAGGATGACGAGGA 20873

RESULT 10

US-09-230-371A-20/c
Sequence 20, Application US/09230371A
Patent No. 6348586
GENERAL INFORMATION:

APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: 45185-G-PCT-US
CURRENT APPLICATION NUMBER: US/09/230,371A
CURRENT FILING DATE: 1999-11-17
PRIOR APPLICATION NUMBER: PCT/US97/13346
PRIOR FILING DATE: 1997-07-22
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 20
LENGTH: 32207
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

Query Match 10.6%; Score 39.6; DB 4; Length 32207;
Best Local Similarity 47.2%; Pred. No. 0.03;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAATCTTTTAAAGACTCTCTGATCTTGACGAATGCAGAAAAACATGTTTCAAGCG 177
DB 21126 GAAGAAGCATCCAGAAATGAAAAAGAAATGTTCCGAAATTAATCAGCTGGCGAGATAT 21067
QY 178 GAGTACTGCTACATCGTTTGAAGACAGGTCAACAAGAAATGTACTACATGTGCTT 237
DB 21066 GGGGACAAACAGATTGACAAAGAAAGTGCAGGTGACAAAGATGACAAATAGAGAT 21007
QY 238 GATGTTGAAGAGTTAGACCAAGAAATTTGTTGCGACGAAACTTCACGAAATTTAT 297
DB 21006 GATGAGGAGAGACAGACAGATGAGAGAGAGATGACGAGAGATGACGAGATGACGAG 20947
QY 298 TTGACAGACTGCGAGGTAAGATGATGATGCGGAGGTACAGGTGACGAGATGACGAGAT 357

Db 20946 GAGGATGACGAGGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAG 20887
QY 358 GAAGTGTGAGAA 371
Db 20886 GAGGATGACGAGGA 20873

RESULT 11

US-08-766-738-2
; Sequence 2, Application US/08766738
; Patent No. 5916749

; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goll, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/766,738
FILING DATE: Herewith

CLASSIFICATION: 530
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0177 US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 966 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:
LIBRARY: Consensus

CLONE: 1813361
US-08-766-738-2

Query Match

Best Local Similarity 49.5%; Score 36; DB 2; Length 966;
Pred. No. 0.077;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 193 GTTTTGAAGACACGCTCAACAAGATGTTACTACATGTCGTTGATGGTGAAGATTA 252
Db 583 GTTGATNNAAGANGAGAGGAGGAGAGAGATGAGAAAGCGAGGAGATGAGAT 642
QY 253 GACCAAGAAAATTTGTTGTCGACGAAAACCTTCAGGAAAATTTATTTGACAGCTGCGAG 312
Db 643 GGTGAAGAGAGAGAGATTTGATGAGAAGATGATGAAGATGAAGATGAAGAGGATGAG 702
QY 313 GGTAAAGATGACAGTAAATGCGGACGATGACAGGTGACAGTCAATGATGAAGAT 372
Db 703 GACGACGATGAAGTCACTGATGAGAGGAGAAAGAAATTTGACTTGATGAAGAGATGAGAT 762
QY 373 TA 374

Db 763 GA 764

RESULT 12

US-09-262-610-2
; Sequence 2, Application US/09262610
; Patent No. 6428949

; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goll, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/262,610
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/766,738
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0177 US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 966 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:
LIBRARY: Consensus

CLONE: 1813361
US-09-262-610-2

Query Match

Best Local Similarity 49.5%; Score 36; DB 4; Length 966;
Pred. No. 0.077;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 193 GTTTTGAAGACACGCTCAACAAGATGTTACTACATGTCGTTGATGGTGAAGATTA 252
Db 583 GTTGATNNAAGANGAGAGGAGGAGAGAGATGAGAAAGCGAGGAGATGAGAT 642
QY 253 GACCAAGAAAATTTGTTGTCGACGAAAACCTTCAGGAAAATTTATTTGACAGCTGCGAG 312
Db 643 GGTGAAGAGAGAGAGATTTGATGAGAAGATGATGAAGATGAAGATGAAGAGGATGAG 702
QY 313 GGTAAAGATGACAGTAAATGCGGACGATGACAGGTGACAGTCAATGATGAAGAT 372
Db 703 GACGACGATGAAGTCACTGATGAGAGGAGAAAGAAATTTGACTTGATGAAGAGATGAGAT 762
QY 373 TA 374
Db 763 GA 764
RESULT 13
US-08-867-941-2

Sequence 2, Application US/08867941
Patent No. 5977337
GENERAL INFORMATION:
APPLICANT: Loomore, Sheena M
APPLICANT: Du, Run-Pan
APPLICANT: Wang, Quljun
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/667,941
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-867-941-2
Query Match 9.3%; Score 35; DB 2; Length 2694;
Best Local Similarity 52.2%; Pred. No. 0.25;
Matches 105; Conservative 0; Mismatches 90; Indels 6; Gaps 1;
QY 172 AAGACGAGTACTGCTACATCGTTTGAAGACACGCTAACAAGAAATGTTACTACAT 231
DB 2017 AAAAAGACAAAGGTTATGCAATATGAAGAAACCATCAAGAAAAAGCCATCAAGAT 2076
QY 232 GTGCTGATGAGTGAAGGTTAGACCAAGAAATTTGTGCGACGAAACTTCACGGA 291
DB 2077 TATCTGTTAACCGAAGACTTCACCCAGAA-----GATGATGACGATGATTTGACCGCA 2130
QY 292 AATTATTGACAGACTGCGAGGTTAAAGATGCAAGTAATCGCGCAGTACAGGTGACGAG 351
DB 2131 TCTGATGATTCACAAGATGATGATGCAAGCGGATGATGATTTGATGATGATGAT 2190
QY 352 TCAGATGAAGTGTGAAGAT 372
DB 2191 TCACAAGATGATGACGAGAT 2211
RESULT 14
US-09-074-658-2
Sequence 2, Application US/09074658
GENERAL INFORMATION:
APPLICANT: Loomore, Sheena M
APPLICANT: Run-Pan Du
APPLICANT: Quljun Wang
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA

NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,658
FILING DATE: 08-MAY-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-795
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1163
TELEFAX: (416) 595-1155
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-658-2
Query Match 9.3%; Score 35; DB 4; Length 2694;
Best Local Similarity 52.2%; Pred. No. 0.25;
Matches 105; Conservative 0; Mismatches 90; Indels 6; Gaps 1;
QY 172 AAGACGAGTACTGCTACATCGTTTGAAGACACGCTAACAAGAAATGTTACTACAT 231
DB 2017 AAAAAGACAAAGGTTATGCAATATGAAGAAACCATCAAGAAAAAGCCATCAAGAT 2076
QY 232 GTGCTGATGAGTGAAGGTTAGACCAAGAAATTTGTGCGACGAAACTTCACGGA 291
DB 2077 TATCTGTTAACCGAAGACTTCACCCAGAA-----GATGATGACGATGATTTGACCGCA 2130
QY 292 AATTATTGACAGACTGCGAGGTTAAAGATGCAAGTAATCGCGCAGTACAGGTGACGAG 351
DB 2131 TCTGATGATTCACAAGATGATGATGCAAGCGGATGATGATTTGATGATGATGAT 2190
QY 352 TCAGATGAAGTGTGAAGAT 372
DB 2191 TCACAAGATGATGACGAGAT 2211

RESULT 15
US-08-867-941-1
Sequence 1, Application US/08867941
Patent No. 5977337
GENERAL INFORMATION:
APPLICANT: Loomore, Sheena M
APPLICANT: Du, Run-Pan
APPLICANT: Wang, Quljun
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,941
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-681 MIS:jfb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1153
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7650 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-867-941-1

Query Match 9.3%; Score 35; DB 2; Length 7650;
Best Local Similarity 52.2%; Pred. No. 0.42;
Matches 105; Conservative 0; Mismatches 90; Indels 6; Gaps 1;
QY 172 AAGACGAGTACTGCTACATGTTTGAAGACACGGTCAACAAGAGATGTTACTACAAT 231
DB 2157 AAAAAAGACAAGGTTATAGCAATATAGAAACATCAAGAAAAAGGCATCAAGAT 2216
QY 232 GTCGTTGATGGTGAAGACTTAGACCAAGAAAAATTTGTTGTCGACGAAAACTTCACGGAA 291
DB 2217 TATCTGTTAACCGAGACTTCACCCAGAA-----GATGATGACGATGATTGACCGCA 2270
QY 292 AATTATTTGACAGACTGGAGGTAAGATGCAGTAATGGCGCAGGTACAGGTGACGAG 351
DB 2271 TCTGATGATTCAACAAGATGATGATGCATGCGCATGATGATTGATTCATCTGATGAT 2330
QY 352 TCAGATGAAGTTGATGAAGAT 372
DB 2331 TCACAAGATGATGACGACGAT 2351

Search completed: May 5, 2003, 17:46:42
Job time : 76 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 16:43:12 ; Search time 138 Seconds
(without alignments)
3210.927 Million cell updates/sec

Title: US-09-936-737A-1
Perfect score: 375
Sequence: 1 atgaagtattcttgattc.....atgaagtatgaagattaa 375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

```
Searched:      746064 segs, 590810554 residues
Total number of hits satisfying chosen parameters: 1492128
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

```
Database : PublishedApplications_NA:*
1: /cgn2_6/ptodata/2/pubnpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubnpna/PCF_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubnpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubnpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubnpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubnpna/PCFUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubnpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubnpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubnpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubnpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubnpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubnpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubnpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubnpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
	1	39.6	10.6	462	9	US-09-918-995-348	Sequence 348, App	
C	2	39.6	10.6	766	10	US-09-864-761-19608	Sequence 19608, App	
C	3	39.6	10.6	1944	10	US-09-864-761-2825	Sequence 2825, App	
	4	37.2	9.9	315	10	US-09-864-761-21723	Sequence 21723, App	
	5	36.6	9.8	499	9	US-09-783-590-5790	Sequence 5790, App	
	6	36.6	9.6	381	9	US-10-015-219-700	Sequence 700, App	
	7	36	9.6	381	10	US-09-777-564-700	Sequence 700, App	
C	8	36	9.6	966	9	US-10-213-700-2	Sequence 2, App1	
C	9	35.8	9.5	204	10	US-09-864-761-21008	Sequence 21008, App	
C	10	35.8	9.5	474	10	US-09-864-761-4255	Sequence 4255, App	
C	11	35.4	9.4	876	10	US-09-770-445-544	Sequence 544, App	
	12	35.4	9.4	1092	9	US-09-938-842A-1125	Sequence 1125, App	
	13	35	9.3	267	10	US-09-878-574-15076	Sequence 15076, App	
	14	35	9.3	520	9	US-10-184-644-332	Sequence 332, App	
	15	35	9.3	520	9	US-10-184-634-332	Sequence 332, App	
	16	35	9.3	563	10	US-09-864-761-13293	Sequence 13293, App	
C	17	34.8	9.3	327	10	US-09-864-761-28059	Sequence 28059, App	
C	18	34.8	9.3	1626	9	US-10-116-016-19	Sequence 19, App1	
	19	34.8	9.3	1626	10	US-09-764-848-19	Sequence 19, App1	

21	34.8	9.3	5361	9	US-09-742-096-2	Sequence 2, Appl 1
20	34.8	9.3	6152	9	US-09-742-096-1	Sequence 1, Appl 1
22	34.6	9.2	438	10	US-09-864-761-4988	Sequence 4988, App
23	34.6	9.2	542	10	US-09-864-761-12239	Sequence 12239, App
24	34.6	9.2	8391	9	US-10-151-736-3	Sequence 3, Appl 1
25	34.6	9.2	8493	9	US-10-151-736-5	Sequence 5, Appl 1
26	34.6	9.2	1651139	9	US-10-067-514-1	Sequence 1, Appl 1
27	34.4	9.2	1846	10	US-09-815-242-8918	Sequence 8918, App
28	34.4	9.2	1282	10	US-10-002-3344-89	Sequence 89, App
29	34.4	9.2	8345	9	US-09-764-891-8451	Sequence 8451, App
30	34.2	9.1	381	10	US-09-864-761-21610	Sequence 21610, App
31	34.2	9.1	659158	9	US-09-771-208-20	Sequence 20, Appl 1
32	33.8	9.0	1824	10	US-09-815-242-9281	Sequence 9281, App
33	33.8	9.0	1824	10	US-09-815-242-9584	Sequence 9584, App
34	33.6	9.0	305	10	US-09-864-761-19262	Sequence 19262, App
35	33.6	9.0	496	10	US-09-864-761-2534	Sequence 2534, App
36	33.4	8.9	420	10	US-09-864-761-23266	Sequence 23266, App
37	33.4	8.9	451	10	US-09-864-761-6554	Sequence 6554, App
38	33.4	8.9	451	10	US-09-864-761-14152	Sequence 14152, App
39	33.4	8.9	7657	10	US-09-070-927A-33	Sequence 33, Appl 1
40	33.2	8.9	157	10	US-09-864-761-29856	Sequence 29856, App
41	33.2	8.9	423	9	US-09-938-842A-592	Sequence 592, App
42	33.2	8.9	58985	9	US-09-901-152-3	Sequence 3, Appl 1
43	33	8.8	432	7	US-08-781-986A-696	Sequence 696, App
44	33	8.8	1959	9	US-09-864-761-4012	Sequence 4012, App
45	33	8.8	2920	10	US-09-801-574-87	Sequence 87, Appl 1

ALIGNMENTS

```

RESULT 1
US-09-918-995-348
: Sequence 348, Application US/09918995
: Publication NO. US20030073623A1
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
: FILE REFERENCE: 20411-756
: FROM VARIOUS CDNA LIBRARIES
: CURRENT APPLICATION NUMBER: US/09/918,995
: PRIOR FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: US/09/235,076
: PRIOR FILING DATE: 1999-01-20
: NUMBER OF SEQ ID NOS: 38054
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 348
: LENGTH: 462
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (1)..(462)
: OTHER INFORMATION: n = A,T,C or G
: US-09-918-995-348

```

Query Match	10.6%	Score 39.6	DB 9	Length 462
Best Local Similarity	50.5%	Pred. No. 0.11		
Matches 96	Conservative 0	Mismatches 94	Indels 0	Gaps 0

[illegible]

LENGTH: 1944
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AI08720.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-2825

Query Match 10.6%; Score 39.6; DB 10; Length 1944;
Best Local Similarity 51.1%; Pred. No. 0.22; Mismatches 89; Indels 0; Gaps 0;
Matches 93; Conservative 0;

QY 187 TACATCGTTTGAAGACACGCTCAACAGGATGTCTACTACATGCTGTGATGCGAA 246
DB 924 TCACACTTGAGCACACCATGAGATGATGTGATGAAGATGATGATGCT 865
QY 247 GAGTTAGACCAAGAAATTTGTTCCACGAAACTTACGGAATTTATTGACAGAC 306
DB 864 GATGAGAGGATGATTAATGATGACATGATGATGATTAATGATGATGAGGAGAG 805
QY 307 TCGAGGGTAAAGATGACAGTATCGGACAGTACAGGTGACGAGTCAGATGAATGAT 366
DB 804 GACGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 745
QY 367 GA 368
DB 744 GA 743

RESULT 4

US-09-864-761-21723
Sequence 21723, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmlca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21723
LENGTH: 315

QY 190 ATCGTTTGAAGACACGCTCAACAGGATGTCTACTACATGCTGTGATGCGAA 249
DB 52 ACGAGGTTTGAGAAACCAAAAAAGAAAGAAAAATCCAGACAGCAAGAACAA 111
QY 250 TTGACCAAGAAATTTGTTGTCACGAAACTTCAAGGAAATTTTTCACAGACTGC 309
DB 112 GAACAAAGAAAG 171
QY 310 GAGGTTAAAGTACAGTATTCGGCGAGTACAGGTGACAGTACAGTACAGTACAGT 369
DB 172 GAAGAAAG 231
QY 370 GATTAA 375
DB 232 GAAGAA 237

Query Match 9.9%; Score 37.2; DB 10; Length 315;
Best Local Similarity 50.0%; Pred. No. 0.43; Mismatches 93; Indels 0; Gaps 0;

Matches 93; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

US-09-783-590-5790
Sequence 5790, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
APPLICANT: Dillon, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16.2C1
CURRENT APPLICATION NUMBER: US/09/783,590
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0

```
SEQ ID NO 5790
LENGTH: 499
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (9)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (127)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (211)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (233)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (262)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (270)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (302)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (320)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (330)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (337)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (345)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (346)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (348)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (368)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (375)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (387)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (388)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (391)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (394)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (395)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (396)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (398)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (408)
```

```
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (417)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (423)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (438)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (439)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (456)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (467)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (468)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (469)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (470)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (476)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (481)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (487)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (492)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (497)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5790
```

Query Match 9.8%; Score 36.6; DB 10; Length 499;
Best Local Similarity 50.6%; Pred. No. 0.82;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
QY 31 CTCGCAAGCTTGCATCTCACTACTCTTCAGAGAACGTGAAGATGTTGGACGTT 90
DB 10 CTAGATTTCTAGTAATCTAAACCCCTCCAGGTACCAAGATCAAGTTGTTGAC 69
QY 91 TACGGACAGCAAAATATATACAGACTTCATTAATCTTTAGAACTCCTCGATCTGAC 150
DB 70 TCCACAGCACACAAAGAGATGTCTTAGATTAAGACAGTAAAGAGATTTATCATAT 129
QY 151 GAATGCAAAAACATGTTTCAAGACGAGTACTCTACATCGTTTGAAG 202
DB 130 CATTTCACTGATTAATCTTTTAAATTGAACATTTGAACATGCTGTGAGG 181
```

RESULT 6
US-10-015-219-700
Sequence 700, Application US/10015219
Publication No. US20030008299A1
GENERAL INFORMATION:
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.493C1
CURRENT APPLICATION NUMBER: US/10/015,219
CURRENT FILING DATE: 2002-03-02

Query Match	9.6%	Score 36	DB 10	Length 381
Best Local Similarity	49.5%	Pred. No. 1.1		
Matches	90	Conservative	0	Mismatches 92
				Indels 0
				Gaps 0
QY	193	GTTTTGAAGACACGGTCACAAAGAAATGTTACTACATGTCGTTGATGTGAAGAGTTA	252	
Db	21	GTGTGATGAAGAGAGAGAGACCAAGAAAGAAAGTGAAGAACNNAGCAGATGAGAT	80	
QY	253	GACCAAGAAAAATTTGTTGTCGACGAAAACCTCACGGAAAAATTTTGCACAGATCGAG	312	
Db	81	GGTGAAGGAAGAGAGAGTTTGATGTCMAAGMAATGATMAAGATGACGATGTGAAGAGCGATGAG	140	
QY	313	GGTAAGATGTCAGGTAAATGGCGGACGTACAGGTACGAGTCACATGAAGTTGATGAAGAT	372	
Db	141	GACGACGATGAAGATCAGTGTAGAGGAGACAAATTTGCACTTGATGTGAAGAAATGAAGAT	200	

QY	193	GTTTTGAAGACACCGCTCAACAAGAGATTGTTACTACAAATGTGTTGATGTGTAAGAGTTA	255	TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
Db	583	GTTTGATTAATGATGAGAGAT	642	TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
QY	253	GACCAAGAAAATTTTGTGTGCAGCAAAACCTTCACGGAAAATTTATTGTGCAGACTGCGAG	312	TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
Db	643	GGTGAAGAAAGAGAGAGATTGATGTAAGAAAGATGATGAAGATGTAAGATGTTAGAAAGGGATGAG	702	TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
QY	313	GGTAAAGATGCAGGTAAATGCGGACAGCTACAGAGTGCAGAGTGCAGATGAAGTTGATGAAGAT	372	TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
Db	703	GACGACGATGAAGTCACTGAGGAGGAGGAAGAATTTTGACTGATTGATGAAGAATGAAAGAT	762	TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
QY	373	TA 374		
Db	763	GA 764		

```

RESULT 9
US-09-864-761-21008/c
Sequence 21008, Application us/09864761
Patient No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21008
LENGTH: 204
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL121580.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HB1100, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.97
OTHER INFORMATION: NT HIT: AF155827.1, EVALUE 1.60e+00
OTHER INFORMATION: EST_HUMAN HIT: BE889896.1, EVALUE 2.80e+00
US-09-864-761-21008

```

```
Query Match          9.5%; Score 35.8; DB 10; Length 204;
Best Local Similarity 54.1%; Pred. No. 0.88;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
```

Oy 238 GATGCTGAAGATTAGACCAAAAAAATTTGTTCGCACGAAAACCTCACCAGAAATTTAT 297
 ||| |||| | | | | | | | | | | | | | | | | |
Db 171 GATGATGAAGTGTGTAATAATCGAATGAMATGACAGAACAAGAATGATTAAATCAA 112

Oy 298 TTGACAGACTCGGAGGGCTAAAGATGCAGCTATCCGCCAGCTACAGGAGCACACTCAT 357
 ||| || | | | | | | | | | | | | | | | | |
Db 111 GATGAAGAAGATGAAAAATGAAGATGAAAAATGAAGTGAGATTAAGATGAAGAAGAA 52

Oy 358 GAAGTTGATGAAGAT 372
 || | | | | | | |
Db 51 GATGAATAATGAAGAT 37

RESULT 10
US-09-864-761-4255/c
Sequence 4255, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 4255
LENGTH: 474
TYPE: DNA

```

Query Match          9.4%; Score35.4; DB 9; Length 1092;
Best Local Similarity 52.3%; Pred. No.2.7;
Matches    78; Conservative    0; Mismatches   71; Indels      0; Gaps      0;

QY     227 ACAATGTCGTGGTGTGAAGAGTTAAGACCAAGAAAAATTTGTTGCAGCAAAACTTCA 286
       ||||| | - - - | - - - ||||| | - - - |
Db      881 ACATGATCTCGAACGACGCTCATCTCGATGAGAAGTAGAGATTGTTGGACTCACAAGTTTA 940

QY     287 CGAAAATAATTATTGACACACTGCCAGGCTAAAGATGACGAGTAATCCGCACGTACAGTG 346
       || | | | | | | | | | | | | | | | | | | | |
Db      941 TTGTGTTAACTCGAAGAGAATGGGTACAGTTGATGAGCAAATATTTTTCTGGATAATGG 1000

QY     347 ACGAGTCAGATGAGTGGTTGTAAGATTAA 375
       | | | | | | | | | | | | | | | |
Db      1001 ATGAGTTCGACGAACCTTGTATCGATGAA 1029

RESULT 13
US-09-878-574-15076
Sequence 15076, Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPPLICATE: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(1540)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
```


GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 16:22:32 ; Search time 1576 Seconds
(without alignments)
3853.622 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 375

Sequence: 1 atgaagattcttctgatttc.....atgaagttgatgaagattaa 375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_fod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51.8	13.8	879	17	AZ550718 ENT56718
2	51.6	13.8	701	13	BM168486 ENT571009
3	51.6	13.8	721	13	BM165474 ENT567997
4	51.2	13.7	881	17	AZ688514 ENTMD22TF
5	50	13.3	236	13	BM163498
6	50	13.3	272	13	BM166089

Result No.	Score	Query Match	Length	DB ID	Description
7	50	13.3	274	13	BM166886 ENT569409
8	50	13.3	276	13	BM163293
9	50	13.3	425	13	BM162507
10	50	13.3	453	13	BM167376
11	50	13.3	459	13	BM171330
12	50	13.3	580	13	BM161175
13	50	13.3	648	13	BM164042
14	50	13.3	651	13	BM166448
15	50	13.3	667	13	BM161293
16	50	13.3	678	13	BM170834
17	50	13.3	702	13	BM169291
18	50	13.3	719	13	BM160442
19	50	13.3	739	13	BM167115
20	50	13.3	753	13	BE846498
21	50	13.3	762	12	BM166415
22	50	13.3	790	13	BM168415
23	49.2	13.1	942	17	BM148582
24	49	13.1	816	17	AZ535744
25	49	13.1	891	17	AZ683582
26	49	13.1	976	17	BM149983
27	48.4	12.9	435	13	BM169710
28	47.4	12.6	900	17	AZ549980
29	47	12.5	843	17	AZ551618
30	47	12.5	877	17	AZ531291
31	47	12.5	908	17	AZ548467
32	47	12.5	912	17	AZ551092
33	47	12.5	931	17	BM160272
34	46.8	12.5	724	13	BM161274
35	46.8	12.5	1006	17	AZ679713
36	46.6	12.4	558	13	BM163267
37	46.6	12.4	890	17	BM146886
38	46.6	12.4	906	17	BM153606
39	46.4	12.4	468	12	BC553005
40	46.4	12.4	880	17	AZ669474
41	46.2	12.3	493	13	BM170539
42	46.2	12.3	1101	17	CNS0181N
43	45.8	12.2	849	17	AZ546009
44	45.8	12.2	905	17	AZ550256
45	45.4	12.1	916	17	AZ671886

ALIGNMENTS

RESULT 1
LOCUS AZ550718/c 879 bp DNA linear GSS 14-NOV-2000
DEFINITION ENT56718 Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ550718
VERSION AZ550718.1 GI:11176019
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 879)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3343
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 25
High quality sequence stop: 801.

FEATURES

Source

Location/Qualifiers
1. 879
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:3759"
/clone_1ib="Entamoeba histolytica Sheared DNA".
/note="Vector: pBOS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of *E. histolytica*
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, U.S. (1993) *Entamoeba histolytica*: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. in Genome
Sequencing: A Practical Approach, eds. M. Vaundin and B.
Barrell, Oxford University Press, 1999)."

BASE COUNT	193 a	219 c	72 g	395 t
ORIGIN				

ORIGIN

Query Match	13.8%;	Score 51.8;	DB 17;	Length 879;
Best Local Similarity	46.9%;	Pred. No. 0.0004;		
Matches 161;	Conservative	0;	Mismatches 182;	Indels 0;
				Gaps 0;

QY	29	GCCTCGACACCTTGATCTCATCTCACTCTCTTGGAAGACGGAAGATGTGGACG	88
Db	768	GACCTGTCTACTGCTTTTAATGCTTTCGTTTCCGAAAGAAAATGAACCTTGACACACAT	709
QY	89	TTTACGGACACAGAAAATATACAGACTTCGATAAATCTTTTAAGAGTCCCTGATCTTG	148
Db	708	TCGATATTGTTTTCACGTAGATTAGATTAACTGACCTTTTGAAGAAAGAAAATATATG	649
QY	149	ACGATTCACAAAAMACATGTTTCAAGACGAGTACTGCTACATGTTTGGACACAG	208
Db	648	AAGTATGTAATTCACATTTAGACGAAGAAGATGATGAAGAAAGATGAAGAATGAAG	589
QY	209	TCAACACGAAATGTTCTACTACATGTCTGTTGATGGTGAAGAGTTGACCACAGAAAATTTG	268
Db	588	AAGTGAAGAAGATGGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATCAACAAATG	529
QY	269	TTGTGCGAGAAAACCTTCACGCAAAATTTATTTTACAGACTCGCAGAGTAAAGATCCAGTA	328
Db	528	AAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGACGAAGAAGATGAAGAAG	469
QY	329	ATGCGCAGATACAGGTGACGAGCTCAGATGAAGATGATGAGA	371
Db	468	ATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAG	426

RESULT 2

BM166486

DEFINITION EST571009 PyBS *Plasmodium yoelii* yoelii cDNA clone PYCP36 5' end, mRNA sequence.

ACCESSION	BM168486	GI:17301718
VERSION	BM168486.1	
KEYWORDS	EST.	

SOURCE ORGANISM	plasmodium	yeell1	yeell1
	plasmodium	yeell1	yeell1

REFERENCE
1 (bases 1 to 701)
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium

AUTHORS
Carlton, J.M., Daly T.M., Long, C.A., Bergman, L.W., Valiyya, A.B.,
Fraser, C.M. and Carucci, D.J.

TITLE Plasmodium yoelii EST project at TIGR
JOURNAL Unpublished (2001)

COMMENT
Contact: Jane Carlton
Parasite Genomics Group

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 984 3818

Tel: 301-530-9319
Fax: 301-838-0208

FEATURES
source

Source

Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ARCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADP.

Location/Qualifiers
1. .701

```

/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCPB36"
/clone_1b="PyB8"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cBYJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size fractionated cDNA was precipitated and ligated to
HybridAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (pad-GAL4) was
excised from the HybridAP vector and plasmid DNA
isolated."

```

BASE COUNT	320 a	62 c	147 g	172 t
ORIGIN				

ORIGIN

Query Match	13.88;	Score 51.6;	DB 13;	Length 701;
Best Local Similarity	55.68;	Pred. No. 0.00044;		
Matches 99;	Conservative 0;	Mismatches 79;	Indels 0;	Gaps 0

Oy 198 TGMGACACGGTCAACAAGSAATTTACTCCAAATCGTGGTGATGCTGAAGAAGTTAAGCCA 257
| | | | | | | | | | | | | | | | | | | | |
Db 512 TGAAAGATGATTAAAGAAGATGAAGATGAAGATGATGACGAAGAAGATBAATGATGTAGACA 571

Oy 258 AGAAAAATTTGTTCGCAGCAAAACTCACGGAATAATTATTGCACACTCGCAGGCTAA 317
| | | | | | | | | | | | | | | | | | | | |
Db 572 GGATCAAAATATACGAAGAGATGTAAGATGACGAAGAAGATGAAGATGACGAAGAAGTGA 631

Oy 318 AGATCAGCATGATCGCGCAGGTACAGTGAAGTACAGTCAATGATGAATTTAA 375
| | | | | | | | | | | | | | | | | | | | |
Db 632 AGATCAACATATATGACAGAGATGAAGATGATGAAGAAGATGAAGAAATGAAGATGA 689

RESULT 3

BM1654/4

DEFINITION EST561997 pyBS plasmodium yoelli yoelli cDNA clone pYCMN68 5' end, mRNA sequence.

ACCESSION	BM165474	GI:17311155
VERSION	BM165474.1	
KEYWORDS	EST.	

SOURCE ORGANISM	Plasmodium yoelii	Plasmodium yoelii
	yoelii	yoelii

REFERENCE
1 (bases 1 to 721)
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium

AUTHORS
Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valoya, A.B.,
Fraser, C.M. and Carucci, D.J.

TITLE Plasmodium yoelli project at TIGR
JOURNAL Unpublished (2001)

COMMENT
Contact: Jane Carlton
Parasite Genomics Group

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 530 0310

Tel: 301-530-9319
Fax: 301-838-0208

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES

source

Location/Qualifiers

1. 236

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="pYCF94"

/clone_11b="PyBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BALB/cByJ mice infected with Py17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was

isolated using the guanidinium isothiocyanate method, and

mRNA isolated using oligo(dT)-cellulose chromatography.

First strand cDNA synthesis was completed using a 50-base

primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven

termini were treated with Pfu DNA polymerase and EcoRI

adaptors ligated to the blunt ends. The sample was cleaved

with XhoI and separated on a Sephacryl S-500 column.

Size-fractionated cDNA was precipitated and ligated to

HybridAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was

excised from the HybridAP vector and plasmid DNA

isolated."

BASE COUNT
ORIGIN

121 a 8 c 68 g 39 t

Query Match

Best Local Similarity 13.3%; Score 50; DB 13; Length 236;
Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Db

198 TGAAGACAGCGTCAACAGAAATGTTACTACATGCTGTGATGATGAGATTAGACCA 257

Db

53 TGAAGATGATGAAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 112

Qy

258 AGAAAAATTTGTTGTCGACGAAAACCTTCACGAAAATTTATTGACAGACTGCGAGGTTAA 317

Db

113 GGATGAAGATGACGAAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 172

Qy

318 AGATGACAGGTATGCGGACGAGTACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 375

Db

173 AGATGAAGATGATGAAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 230

RESULT 6

BM166089 272 bp mRNA linear EST 04-DEC-2001
LOCUS BM166089
DEFINITION EST568612 PyBS Plasmodium yoelii yoelii cDNA clone pYCN52 5' end,
mRNA sequence.

ACCESSION

BM166089

VERSION

BM166089.1 GI:17299321

KEYWORDS

EST.

SOURCE

Plasmodium yoelii yoelii.

ORGANISM

Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE

1 (bases 1 to 272)
Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,
Fraser, C.M. and Carucci, D.J.

AUTHORS

Plasmodium yoelii EST project at TIGR

TITLE

Plasmodium yoelii EST project at TIGR

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES

source

Location/Qualifiers

1. 272

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="pYCN52"

/clone_11b="PyBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BALB/cByJ mice infected with Py17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was

isolated using the guanidinium isothiocyanate method, and

mRNA isolated using oligo(dT)-cellulose chromatography.

First strand cDNA synthesis was completed using a 50-base

primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven

termini were treated with Pfu DNA polymerase and EcoRI

adaptors ligated to the blunt ends. The sample was cleaved

with XhoI and separated on a Sephacryl S-500 column.

Size-fractionated cDNA was precipitated and ligated to

HybridAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was

excised from the HybridAP vector and plasmid DNA

isolated."

BASE COUNT
ORIGIN

137 a 11 c 75 g 49 t

Query Match

Best Local Similarity 13.3%; Score 50; DB 13; Length 272;
Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Db

198 TGAAGACAGCGTCAACAGAAATGTTACTACATGCTGTGATGATGAGATTAGACCA 257

Db

89 TGAAGATGATGAAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 148

Qy

258 AGAAAAATTTGTTGTCGACGAAAACCTTCACGAAAATTTATTGACAGACTGCGAGGTTAA 317

Db

149 GGATGAAGATGACGAAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 208

Qy

318 AGATGACAGTATGCGGACGAGTACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 375

Db

209 AGATGAAGATGATGAAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 266

RESULT 7

BM166886 274 bp mRNA linear EST 04-DEC-2001
LOCUS BM166886
DEFINITION EST569409 PyBS Plasmodium yoelii yoelii cDNA clone pYCN214 5' end,
mRNA sequence.

ACCESSION

BM166886

VERSION

BM166886.1 GI:17300118

KEYWORDS

EST.

SOURCE

Plasmodium yoelii yoelii.

ORGANISM

Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE

1 (bases 1 to 274)
Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,
Fraser, C.M. and Carucci, D.J.

AUTHORS

Plasmodium yoelii EST project at TIGR

TITLE

Plasmodium yoelii EST project at TIGR

Unpublished (2001)
Contact: Jane Carlton

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
5712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-8919
Fax: 301-838-0208

Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ARCC
<http://www.malaria.rtt4.org/mr4pages/index.html>
Seq primer: ADP.

Location/Qualifiers
1. .453

/organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db.xref="taxon:73239"
 /clone="PYC0128"
 /clone.lib="PYBS"
 /dev.stage="Asexual blood stages"
 /lab.host="E. coli XL-1 Blue"
 /note="Vector: pBD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cBYJ mice infected with PY17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase, and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HygriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pBD-GAL4) was excised from the HygriZAP vector and plasmid DNA isolated."

223 a 25 c 112 g 93 t

Query Match	13.3%	Score 50;	DB 13;	Length 453;
Best Local Similarity	55.1%;	Pred. No. 0.0012;		
Matches	98;	Conservative	0;	Mismatches 80;
			Indels	0;
			Gaps	0;

[illegible]

LOCUS	BM171330	459 bp	mRNA	linear	EST 04-DEC-2001
DEFINITION	EST573853 pyBS plasmodium yoelli yoelli cDNA clone pYCqT11 5' end, mRNA sequence.				

BM171330
BM171330.1 GI:17304562

KEYWORDS EST. *Plasmodium yoelii yoelii*
SOURCE

Plasmodium yoelii yoelii
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 459)

Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdy, A.B.,

Plasmodium yoelii EST project at TIGR

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES
source

Location/Qualifiers
1..459
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYC011"
/clone_1lb="PYC011"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

BASE COUNT 225 a 37 c 113 g 84 t
ORIGIN

Query Match 13.3%; Score 50; DB 13; Length 459;
Best Local Similarity 55.1%; Pred. No. 0.0012;
Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAAGACACGGTCAACAGGATGTACTACATCTCGTTGCTGAAGCTTGAACCA 257
|||||
DB 108 TGAAGATGATGAGAGATGAAGATGATGACGAAAGATGAAGATGATGAGA 167
QY 238 AGAAAAATTTGTTGTCGACGAAACTTCACGGAATTTATTTGACAGACTCGAGGGTAA 317
|||||
DB 168 GGATGAGATGACGAAAGATGAAGATGACGAAAGATGAAGATGAAGATGA 227
QY 318 AGATGCAGTATGCGGACGAGTACAGTACAGTACAGTACAGTACAGTACAGTAA 375
|||||
DB 228 AGATGAAGATGATGAAGAGATGAAGATGATGAAGAGATGAAGAGATGAAGATGAA 285
|||||

RESULT 12
BM161175 580 bp mRNA linear EST 04-DEC-2001
LOCUS BM161175
DEFINITION EST563698 PYBS Plasmodium yoelii yoelii cDNA clone PYCRB54 5' end,
mRNA sequence.
ACCESSION BM161175
VERSION BM161175.1 GI:17306856
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
REFERENCE 1 (bases 1 to 580)
AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,
Fraser,C.M. and Carucci,D.J.
TITLE Plasmodium yoelii EST project at TIGR

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES
source

Location/Qualifiers
1..580
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCRB54"
/clone_1lb="PYCRB54"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

BASE COUNT 251 a 47 c 139 g 143 t
ORIGIN

Query Match 13.3%; Score 50; DB 13; Length 580;
Best Local Similarity 55.1%; Pred. No. 0.0012;
Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAAGACACGGTCAACAGGATGTACTACATCTCGTTGCTGAAGCTTGAACCA 257
|||||
DB 397 TGAAGATGATGAGAGATGAAGATGATGACGAAAGATGAAGATGATGAGA 456
QY 258 AGAAAAATTTGTTGTCGACGAAACTTCACGGAATTTATTTGACAGACTCGAGGGTAA 317
|||||
DB 457 GGATGAGATGACGAAAGATGAAGATGACGAAAGATGAAGATGAAGATGA 516
QY 318 AGATGCAGTATGCGGACGAGTACAGTACAGTACAGTACAGTACAGTACAGTAA 375
|||||
DB 517 AGATGAAGATGATGAAGAGATGAAGATGATGAAGAGATGAAGAGATGAAGATGAA 574
|||||

RESULT 13
BM164042 648 bp mRNA linear EST 04-DEC-2001
LOCUS BM164042
DEFINITION EST565655 PYBS Plasmodium yoelii yoelii cDNA clone PYCLP02 5' end,
mRNA sequence.
ACCESSION BM164042
VERSION BM164042.1 GI:17309723
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
REFERENCE 1 (bases 1 to 648)
AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,
Fraser,C.M. and Carucci,D.J.
TITLE Plasmodium yoelii EST project at TIGR

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADP.

FEATURES

Source

Location/Qualifiers

1..648

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCNP02"

/clone_1lb="PyBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BALB/cBYJ mice infected with PY17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was

isolated using the guanidium isothiocyanate method, and

mRNA isolated using oligo(dT)-cellulose chromatography.

First strand cDNA synthesis was completed using a 50-base

primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven

termini were treated with Pfu DNA polymerase and EcoRI

adaptors ligated to the blunt ends. The sample was cleaved

with XhoI and separated on a Sephacryl S-500 column.

Size-fractionated cDNA was precipitated and ligated to

HybridZAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was

excised from the HybridZAP vector and plasmid DNA

isolated."

BASE COUNT 319 a 37 c 133 g 159 t
ORIGIN

Query Match

13.3%; Score 50; DB 13; Length 648;

Best Local Similarity 55.1%; Pred. No. 0.0012;

Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAAGACACGGTCAACAGATGTTACTACATGCTGTTGATGTAAGATTAGACCA 257

DB 349 TGAAGATGATGAAGAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 408

QY 258 AGAAAATTTTGTGTCGACGAAACTTCACGGAATTTATTGACAGACTCGAGGGTAA 317

DB 409 GGATGAAGATGACGAGAGAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 468

QY 318 AGATGAGTAAATGCGGACGAGTACGAGTACGAGTACGAGTACGAGTACGAGTAA 375

DB 469 AGATGAAGATGATGAAGAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 526

RESULT 14

BM166448

LOCUS

DEFINITION 651 bp mRNA linear EST 04-DEC-2001

EST568971 PYBS Plasmodium yoelii yoelii cDNA clone pYCNL11 5' end,

mRNA sequence.

ACCESSION BM166448

VERSION BM166448.1 GI:17299680

KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii.

ORGANISM Plasmodium yoelii yoelii.

REFERENCE 1 (bases 1 to 651)

AUTHORS Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,

Fraser, C.M., and Carucci, D.J.

TITLE Plasmodium yoelii EST project at TIGR

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADP.

FEATURES

Source

Location/Qualifiers

1..651

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCNL11"

/clone_1lb="PyBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BALB/cBYJ mice infected with PY17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was

isolated using the guanidium isothiocyanate method, and

mRNA isolated using oligo(dT)-cellulose chromatography.

First strand cDNA synthesis was completed using a 50-base

primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven

termini were treated with Pfu DNA polymerase and EcoRI

adaptors ligated to the blunt ends. The sample was cleaved

with XhoI and separated on a Sephacryl S-500 column.

Size-fractionated cDNA was precipitated and ligated to

HybridZAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was

excised from the HybridZAP vector and plasmid DNA

isolated."

BASE COUNT 291 a 56 c 145 g 159 t
ORIGIN

Query Match

13.3%; Score 50; DB 13; Length 651;

Best Local Similarity 55.1%; Pred. No. 0.0012;

Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAAGACACGGTCAACAGATGTTACTACATGCTGTTGATGTAAGATTAGACCA 257

DB 468 TGAAGATGATGAAGAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 527

QY 258 AGAAAATTTTGTGTCGACGAAACTTCACGGAATTTATTGACAGACTCGAGGGTAA 317

DB 528 GGATGAAGATGACGAGAGAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 587

QY 318 AGATGAGTAAATGCGGACGAGTACGAGTACGAGTACGAGTACGAGTACGAGTAA 375

DB 568 AGATGAAGATGATGAAGAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 645

RESULT 15

BM161293

LOCUS

DEFINITION 667 bp mRNA linear EST 04-DEC-2001

EST56816 PYBS Plasmodium yoelii yoelii cDNA clone pYCN09 5' end,

mRNA sequence.

ACCESSION BM161293

VERSION BM161293.1 GI:17306974

KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii.

ORGANISM Plasmodium yoelii yoelii.

REFERENCE 1 (bases 1 to 667)

AUTHORS Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,

Fraser, C.M., and Carucci, D.J.

TITLE Plasmodium yoelii EST project at TIGR

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2003, 17:29:38 ; Search time 56 Seconds
(without alignments)
245.086 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573
Sequence: 1 EERDCWTFYANRKYTFEDK.....GKDAGNAGTGDSEDEVDED 103

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	573	100.0	103	21	AA94746
2	80.5	14.0	147	13	AA22950
3	80.5	14.0	412	20	AA68010
4	77	13.4	368	22	AB69867
5	75.5	13.2	418	21	AA14260
6	75.5	13.2	419	20	AA50109
7	75.5	13.2	419	21	AA14259
8	75.5	13.2	419	21	AA29108
9	75.5	13.2	432	16	AA66769
10	75.5	13.2	432	17	AA98464

11	73.5	12.8	1154	22	AB66916	Drosophila melanog
12	71.5	12.5	175	21	AA17789	Arabidopsis thalia
13	71.5	12.5	256	21	AA17788	Arabidopsis thalia
14	71.5	12.5	320	21	AA18263	Plasmodium falcipa
15	71	12.4	589	22	AA695086	Human protein sequ
16	70.5	12.3	156	23	AA69411	Lung small cell ca
17	70.5	12.3	367	22	AB71954	Drosophila melanog
18	70	12.2	669	23	AB69771	Amino acid sequenc
19	70	12.2	669	23	AB69781	Amino acid sequenc
20	70	12.2	1167	22	AA035929	Helicobacter pylor
21	69.5	12.1	674	19	AA661238	Streptococcus pneu
22	69.5	12.1	674	23	AB54656	S. pneumoniae Sp11
23	69.5	12.1	3257	22	AB67502	Drosophila melanog
24	69	12.0	717	21	AA641819	Human ORF1583
25	69	12.0	717	22	AA639298	Human polypeptide
26	69	12.0	734	22	AB11943	Human dJ47587.2 ho
27	69	12.0	734	22	AA692929	Human polypeptide
28	69	12.0	738	22	AA694096	Human protein sequ
29	69	12.0	738	22	AA641084	Human polypeptide
30	69	12.0	738	22	AA641085	Human polypeptide
31	68.5	12.0	610	23	AB661879	Prostate cancer-as
32	68.5	12.0	635	21	AA618176	Plasmodium falcipa
33	68.5	12.0	1869	22	AB58651	Drosophila melanog
34	68	11.9	146	23	AB608153	Murine PLA2 enzyme
35	68	11.9	403	22	AA694657	Human protein sequ
36	68	11.9	561	20	AA41665	Zea mays AAC-11 pr
37	68	11.9	867	22	AB607270	Novel human diagno
38	68	11.9	903	22	AA639516	Human polypeptide
39	68	11.9	961	22	AA678435	Human protein S80
40	67.5	11.8	117	23	AB600862	Human ORF1 protein
41	67.5	11.8	366	21	AA653207	Arabidopsis thalia
42	67.5	11.8	450	21	AA634636	Arabidopsis thalia
43	67.5	11.8	450	21	AA653206	Arabidopsis thalia
44	67.5	11.8	488	21	AA634635	Arabidopsis thalia
45	67.5	11.8	488	21	AA653205	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AA94746
ID AA94746 standard; Protein: 103 AA.
XX
AC AA94746;
XX
DT 29-JAN-2001 (first entry)
XX
DE Platelet binding inhibitor protein Saratin amino acid sequence.
XX
KW Saratin; medicinal leech; thromboembolic disease; intraocular lens;
KW collagen-dependent platelet adhesion inhibition;
KW posterior capsule opacification.
XX
XX Hirudo medicinalis.
OS
XX
PN WO200056885-A1.
XX
PD 28-SEP-2000.
XX
PF 10-MAR-2000; 2000WO-EP02117.
XX
PR 18-MAR-1999; 99EP-0105530.
XX
PR 12-MAY-1999; 99EP-0109503.
XX
XX (MERE) MERCK PATENT GMBH.
XX
PI Strittmatter W, Guessow D, Hofmann U, Hemberger J, Fotev Z;
PI Scheuble B;
XX
DR WPI: 2000-611629/58.
XX
DR N-PDB: AA28180.
XX

PT New Saratin polypeptide and gene isolated from Hirudo medicinalis for
 PT blocking platelet adhesion, especially useful for treating or
 PT preventing thrombotic diseases, or for manufacturing a medicament for
 PT thromboembolic diseases

PS Claim 7; Page 41; 46pp; English.

CC Saratin is a protein isolated from the saliva of the medicinal leech
 CC Hirudo medicinalis. Saratin is an inhibitor of collagen-dependent
 CC platelet adhesion. The invention includes polynucleotide sequences
 CC encoding Saratin, an expression vector comprising the DNA sequence, a
 CC host cell transformed with the expression vector, antibodies
 CC immunospecific for Saratin, and methods for identifying Saratin agonists
 CC or antagonists. Saratin is useful for treating thromboembolic diseases,
 CC and for manufacturing a medicament for treating thromboembolic diseases.
 CC It is useful for preventing thrombotic diseases. Saratin is also useful
 CC for coating artificial surfaces, since use of Saratin renders them
 CC non-adhesive for cells and prevents the activation of cells. It may also
 CC be used for coating natural collagen surfaces. Furthermore, Saratin is
 CC useful for modifying intracellular lenses in order to lessen the
 CC thrombogenicity of the lens material, for contacting the lens surface, or
 CC for covalent crosslinking to modify the lens material. The lens material
 CC is used for refractive anterior or posterior chamber ocular implants,
 CC which may be implanted into the eye. This new type of coating avoids
 CC problems contributed by stimulated cell growth. In combination with other
 CC medicaments that are for instance conferring cell death, Saratin coating
 CC helps to completely overcome posterior capsule opacification. The
 CC antibody immunospecific for Saratin, as well as Saratin itself, are
 CC useful for measuring samples derived from host cell cultures or from a
 CC treated subject. The present sequence represents the amino acid sequence
 CC of the Saratin protein.

SQ Sequence 103 AA;

Query Match 100.0%; Score 573; DB 21; Length 103;

Best Local Similarity 100.0%; Pred. No. 2.3e-52;

Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEREDCMTFYANKRYTDFDKSFKSSDLDCKCTCKTECYIVFEDTVNKECYVAVDG 60

DB 1 EEREDCMTFYANKRYTDFDKSFKSSDLDCKCTCKTECYIVFEDTVNKECYVAVDG 60

QY 61 EELDQEKFYVDENFTENYLTDCSGKADGNAAGTDESDEVDED 103

DB 61 EELDQEKFYVDENFTENYLTDCSGKADGNAAGTDESDEVDED 103

RESULT 2

AAR22950 ID AAR22950 standard; Protein; 147 AA.

AC AAR22950;

DT 10-SEP-1992 (first entry)

DE Leech antiplatelet protein.

XX LAPP; collagen; platelet aggregation; antithrombotic; anticoagulant;

KW coronary artery disease; cerebrovascular disease; ss.

XX Haementeria officinalis.

XX Key Location/Qualifiers

FT Peptide 1..21

FT Protein /note="21 hydrophobic amino acids"

XX EP480651-A.

PD 15-APR-1992.

PF 07-OCT-1991; 91EP-0309157.

XX 09-OCT-1990; 90US-0594917.

XX (MERI) MERCK & CO INC.

XX Connolly TN, Keller PW;

DR WPI: 1992-125288/16.

DR N-PSDB; AAQ23682.

PS Claim 6; Page 15; 20pp; English.

CC When cloned into the yeast expression vector pKHA2, the LAPP gene

CC sequence forms a fusion product of MAF-alpha-1 leader sequence

CC plus mature LAPP, which when produced are proteolytically processed by

CC the Lys-Arg cleaving endopeptidase (KE2), which cleaves on the

CC C terminal side of Lys-Arg, and the products secreted into the culture

CC medium.

CC LAPP blocks stimulation of platelet aggregation by collagen, at an

CC antithrombotic concentration of about 100 nMolar (or 1.6 ug/ml).

CC Platelet aggregation is stimulated by 2 ug/ml collagen and addition

CC of LAPP inhibits this with an IC50 of 45nM, but this may be

CC overcome by the addition of 0.25 mM ascorbic acid.

CC LAPP could be used for treatment and prevention of thrombotic

CC conditions especially coronary artery and cerebrovascular disease

CC in mammals.

CC See also AAQ23871-7.

SQ Sequence 147 AA;

Query Match 14.0%; Score 80.5; DB 13; Length 147;

Best Local Similarity 32.7%; Pred. No. 1.1;

Matches 18; Conservative 10; Mismatches 24; Indels 3; Gaps 1;

QY 5 DCMTFYANKRYTDFDKSFKSSDLDCKCTCKTECYIVFEDTVNKECYVAVDG 56

DB 66 DCMSKRPGMKLPDNLITKTEFTSVDECKRMCEASVVERSCYIIQINTFTNCCYRN 120

RESULT 3

AAW68010 ID AAW68010 standard; Protein; 412 AA.

AC AAW68010;

DT 27-APR-1999 (first entry)

DE S. frugiperda immunophilin FRBP46.

XX Immunophilin; moth; insect cell; nuclear; immunosuppression; drug;

KW Transplant; tissue graft.

XX Spodoptera frugiperda.

XX Key Location/Qualifiers

FT Domain 111..114

FT Region /note="EAP motif"

FT Region /note="putative nuclear localisation signal"

FT Region /note="putative nuclear localisation signal"

FT Domain /note="putative nuclear localisation signal"

FT Domain /note="AP motif"

FT Domain /note="EAP motif"

FT Region /note="putative nuclear localisation signal"


```

FT Region 250..256
FT /note= "putative nuclear localisation signal"
FT Region 272..275
FT /note= "putative nuclear localisation signal"
FT Region 284..286
FT /note= "putative nuclear localisation signal"
FT Domain 292..295
FT /note= "EEAP motif"
FT Region 298..302
FT /note= "putative nuclear localisation signal"
XX
XX US5861498-A.
XX
XX 19-JAN-1999.
XX
XX 31-OCT-1996; 96US-0741134.
XX
XX 01-NOV-1995; 95US-0007163.
XX 31-OCT-1996; 96US-0741134.
XX
XX (UYJE-) UNIV JEFFERSON THOMAS.
XX
XX Alnemri ES, Fernandes-Alnemri T, Litwack G;
XX WPI; 1999-130433/11.
XX N-PSDB; AAX04441.
XX
XX Isolated immunophilin FKBP46 nucleic acids - useful for developing
XX products for the study and identification of immunosuppressive
XX agents for treating e.g. transplant and tissue graft patients
XX
XX Claim 1; Fig 1; 21pp; English.
XX
XX This sequence represents the immunophilin FKBP46 from the moth Spodoptera
XX frugiperda. FKBP46 is a novel insect cell nuclear immunophilin, which
XX can be used to study and identify additional immunosuppressive drugs that
XX bind to it. Such immunosuppressive agents can be used in treating
XX transplant and tissue graft patients.
XX
XX Sequence 412 AA;
XX
XX Query Match 14.0%; Score 80.5; DB 20; Length 412;
XX Best Local Similarity 30.6%; Pred. No. 3.9;
XX Matches 30; Conservative 12; Mismatches 29; Indels 27; Gaps 5;
XX
XX QY 11 ANRKYTDPEKSFKKS-----SDDECKKTKCFKTEYCYIVE--DYVKNKCYNVVDGEELDQ 65
XX 133 ANKRAKPDKAGKNSAPAESDSDDD-----EDQLQK-----FLDGEDIDR 174
XX
XX QY 66 KEFVVDENFTENYLTDEGKDGAGNAAGTGDSEDEVED 103
XX 175 DE--NDESFKN--TSAEGDSDSEDEDEDEDEDD 208
XX
XX RESULT 4
XX ABB69867
XX ID ABB69867 standard; Protein; 368 AA.
XX AC ABB69867;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 36393.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO2001/71042-A2.
XX
XX 27-SEP-2001.
XX

```

```

PF 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL13970.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 36393; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 368 AA;
XX
XX Query Match 13.4%; Score 77; DB 22; Length 368;
XX Best Local Similarity 23.9%; Pred. No. 7.8;
XX Matches 28; Conservative 21; Mismatches 34; Indels 34; Gaps 6;
XX
XX QY 14 KYTDFPKS-----FKSSDDECKKTKCFKTEYCYIVE--DYVKNKCYNVVDG 60
XX 89 EYTDADSEVSAPTEFLAEFLSAVMLKDYKKA--LKYCKLIIQYEPDNATAEFPPLI-- 143
XX
XX Db 61 EELDEKFEVY-----DENFTEN-----YLTDEGKDGAGNAAGTGDSEDEVED 103
XX 144 --LDRKRAVATSDSDSENNKSSPDLALDLHASDVADVGDDEAGDADDEGDADAD 198
XX
XX RESULT 5
XX AAB14260
XX ID AAB14260 standard; Protein; 418 AA.
XX AC AAB14260;
XX
XX 09-FEB-2001 (first entry)
XX
XX Mouse Ice-4 protein sequence #2.
XX
XX ced-3; virally induced cell death; apoptosis; gene therapy; neural;
XX muscular degenerative disease; myocardial infarction; stroke; aging;
XX interleukin-beta converting enzyme; ICE; cysteine protease; mouse.
XX
XX Mus sp.
XX
XX US6083735-A.
XX
XX 04-JUL-2000.
XX
XX 10-JUN-1994; 94US-0258287.
XX
XX 24-JUN-1993; 93US-0080850.
XX
XX (GEHO ) GEN HOSPITAL CORP.
XX
XX Yuan J, Miura M;
XX

```

DR WPI; 2000-464343/40.
 XX
 PT New human Ich-1L and Ich-1S proteins for negative and positive
 PT regulation of programmed cell death and for developing therapeutic
 PT methods for diseases and conditions characterized by cell death, e.g.
 PT myocardial infarction or stroke
 PS
 PS Disclosure; Fig 17; 121pp; English.
 XX
 CC The present sequence is a mouse Ice-4 protein sequence. The present
 CC sequence was used in a sequence homology comparison with the protein
 CC sequences of human Ice-4 (Ich-1) (AAB14253), murine
 CC Interleukin-1beta converting enzyme (MICE) (AAB14249), murine ICE2
 CC (AAB14252) and C. elegans ced-3 (AAB14246). The coding sequence of the
 CC present sequence is a member of a family of genes involved in programmed
 CC cell death (apoptosis). Ich-1 may play an important role in both the
 CC positive and negative regulation of apoptosis. The Ich gene may be used
 CC in gene therapy in disorders characterized by cell death e.g. neural and
 CC muscular degenerative diseases, myocardial infarction, stroke, virally
 CC induced cell death and aging.
 CC
 SQ Sequence 418 AA;
 Query Match 13.2%; Score 75.5; DB 21; Length 418;
 Best Local Similarity 35.6%; Pred. No. 13;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;
 Oy 44 VFEDTVNKECYNYVDGEEL---DQEKFYVD--ENFTENYL--TDCCKG-DAGNAGTG 94
 Db 26 VFDDLVK-----NVNDELKRTIGESASFLNKAKENLVENLEKTDMAKIRAGHIANSQ 81
 Oy 95 DE-----SDEVD 101
 Db 82 EQLSLQFSNDEDD 94
 RESULT 6
 ID AAY50109 standard; Protein; 419 AA.
 AC AAY50109;
 XX
 XX 21-JAN-2000 (first entry)
 DE Murine caspase-12 splice variant, caspase-12L.
 XX
 KW Caspase; splice variant; truncated; programmed cell death; apoptosis;
 KW regulation; proteolytic cascade; malignant condition; cancer;
 KW premalignant condition; solid tumour; lymphoma;
 KW chronic lymphocytic leukaemia; prostatic hypertrophy;
 KW preneoplastic liver focus; chemotherapy resistance; autoimmune disease;
 KW ds.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Cleavage-site 94..95
 FT Protein 95..419
 FT /note="Caspase-12"
 FT Cleavage-site 318..319
 XX
 XX W0952925-A1.
 XX
 XX 21-OCT-1999.
 XX
 XX 14-APR-1999; 99MO-US08064.
 XX
 XX 16-APR-1998; 9805-0081962.
 XX
 XX (GEHO) GEN HOSPITAL CORP.
 XX
 XX Yuan J, Morishima N;
 XX
 DR

DR WPI; 1999-620369/53.
 DR N-PSDB; AA32645.
 XX
 PT New nucleic acid encoding the short form of caspase-12, used e.g. for
 PT treating tumors
 PS
 PS Claim 1; Fig 1; 68pp; English.
 XX
 CC This sequence represents a cDNA encoding murine caspase-12L, a splice
 CC variant of caspase-12. Caspases are a family of proteins involved in
 CC the regulation of apoptosis and are synthesised as proforms which are
 CC activated via cleavage after specific Asp residues. Mammalian cells
 CC express several caspases, and it is thought that these act in a
 CC proteolytic cascade to cause programmed cell death. Nucleic acids
 CC encoding caspase-12S (AA32644) or truncated forms of caspase-12L
 CC (AA32646, AA32647) are used for production, recombinantly or in vivo,
 CC of caspase-12 polypeptides which induce programmed cell death. This is
 CC particularly useful for treating (pre)malignant conditions (e.g., solid
 CC tumours, B cell lymphoma, chronic lymphocytic leukaemia, prostatic
 CC hypertrophy, preneoplastic liver foci and resistance to chemotherapy), or
 CC autoimmune diseases. The caspase-12 proteins can also be used to raise
 CC specific antibodies (for example, to determine gene expression and to
 CC screen expression libraries) or as molecular weight markers. Fragments
 CC of caspase-12 encoding nucleic acids can be used as probes to isolate
 CC the caspase-12 gene (and its allelic variants); in fluorescent in situ
 CC hybridisation for chromosomal location of the caspase-12 gene, and for
 CC Northern blotting to determine caspase-12 mRNA expression in tissues.
 CC
 SQ Sequence 419 AA;
 Query Match 13.2%; Score 75.5; DB 20; Length 419;
 Best Local Similarity 35.6%; Pred. No. 13;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;
 Oy 44 VFEDTVNKECYNYVDGEEL---DQEKFYVD--ENFTENYL--TDCCKG-DAGNAGTG 94
 Db 26 VFDDLVK-----NVNDELKRTIGESASFLNKAKENLVENLEKTDMAKIRAGHIANSQ 81
 Oy 95 DE-----SDEVD 101
 Db 82 EQLSLQFSNDEDD 94
 RESULT 7
 ID AAB14259 standard; Protein; 419 AA.
 AC AAB14259;
 XX
 XX 09-FEB-2001 (first entry)
 DE Mouse Ice-4 protein sequence #1.
 XX
 KW ced-3; virally induced cell death; apoptosis; gene therapy; neural;
 KW muscular degenerative disease; myocardial infarction; stroke; aging;
 KW Interleukin-1beta converting enzyme; ICE; cysteine protease; mouse.
 XX
 OS Mus sp.
 XX
 XX US6083735-A.
 XX
 XX 04-JUL-2000.
 XX
 XX 10-JUN-1994; 94US-0258287.
 XX
 XX 24-JUN-1993; 93US-0080850.
 XX
 XX (GEHO) GEN HOSPITAL CORP.
 XX
 XX Yuan J, Miura M;
 XX
 XX WPI; 2000-464343/40.
 XX N-PSDB; AAA72843.
 DR

XX New human Ich-1L and Ich-1S proteins for negative and positive
 PT regulation of programmed cell death and for developing therapeutic
 PT methods for diseases and conditions characterized by cell death, e.g.
 PT myocardial infarction or stroke
 PS Disclosure; Fig 16; 121pp; English.
 CC The present sequence is a mouse Ice-4 protein sequence. The present
 CC sequence was used in a sequence homology comparison with the protein
 CC sequences of human Ice-3 homolog (Ich-1) (AAB14253), murine
 CC interleukin-1beta converting enzyme (mICE) (AAB14249), murine ICE2
 CC (AAB1452) and C. elegans ced-3 (AAB14246). The coding sequence of the
 CC present sequence is a member of a family of genes involved in programmed
 CC cell death (apoptosis). Ich-1 may play an important role in both the
 CC positive and negative regulation of apoptosis. The Ich gene may be used
 CC in gene therapy in disorders characterized by cell death e.g. neural and
 CC muscular degenerative diseases, myocardial infarction, stroke, virally
 CC induced cell death and aging.
 SQ Sequence 419 AA;
 Query Match 13.2%; Score 75.5; DB 21; Length 419;
 Best Local Similarity 35.6%; Pred. No. 13;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;
 OY 44 VEDIVNRECYNVVDGEEL---DQEFVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
 DB 26 VEDDLVER---NVLNGDELKIGESAFILNKAEVLNENLEKTDAGKIFAGHIANSO 81
 OY 95 DE-----SDEVD 101
 DB 82 EQLSIQFSNDEDD 94
 RESULT 8
 AAB29108
 ID AAB29108 standard; protein; 419 AA.
 XX AAB29108;
 AC 07-FEB-2001 (first entry)
 DT 07-FEB-2001 (first entry)
 DE Murine caspase-12 protein.
 XX
 KW Mouse; caspase-12; calpain; apoptosis; cell death; autoimmune disease;
 KW cancer; viral infection; apoptotic condition; gene therapy.
 XX
 OS Mus musculus.
 XX
 PN WO200059924-A1.
 XX
 PD 12-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US09173.
 XX
 PR 06-APR-1999; 99US-0127967.
 XX
 PA (HARD) HARVARD COLLEGE.
 XX
 PI Yuan J, Nakagawa T;
 XX
 DR WPI: 2000-679352/66.
 XX
 PT Caspase-12 polypeptide useful for treating apoptotic conditions e.g.
 PT stroke, ischemia and proliferated disease e.g. cancer, lacks at least 5
 PT amino acids in a specific region of the native amino acid sequence -
 XX
 PS Claim 1; Page 65-66; 70pp; English.
 XX
 CC The present invention is concerned with methods for identifying compounds
 CC useful for modulating caspase-mediated apoptosis. These include the
 CC caspase-12 protein shown here. These agents can be used in the treatment

CC of apoptotic conditions, including stroke, neurodegeneration, ischemia,
 CC Alzheimer's disease and muscular dystrophy, and proliferative diseases
 CC such as cancer, autoimmune disorders such as systemic lupus erythematosus
 CC and multiple sclerosis, and viral infections, particularly those
 CC associated with herpes virus, poxvirus and adenovirus. The treatment may
 CC be in the form of gene therapy.
 SQ Sequence 419 AA;
 Query Match 13.2%; Score 75.5; DB 21; Length 419;
 Best Local Similarity 35.6%; Pred. No. 13;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;
 OY 44 VEDIVNRECYNVVDGEEL---DQEFVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
 DB 26 VEDDLVER---NVLNGDELKIGESAFILNKAEVLNENLEKTDAGKIFAGHIANSO 81
 OY 95 DE-----SDEVD 101
 DB 82 EQLSIQFSNDEDD 94
 RESULT 9
 AAR66769
 ID AAR66769 standard; protein; 432 AA.
 XX AAR66769;
 AC 12-SEP-1995 (first entry)
 DT 12-SEP-1995 (first entry)
 DE Human interleukin-1 beta converting enzyme homolog (Ice-4).
 XX
 KW Human interleukin-1 beta converting enzyme homolog; Ice-4;
 KW oncogene bcl-2; programmed cell death; cancer treatment.
 XX
 OS Homo sapiens.
 XX
 PN WO9500160-A.
 XX
 PD 05-JAN-1995.
 XX
 PF 10-JUN-1994; 94WO-US06630.
 XX
 PR 24-JUN-1993; 93US-0080850.
 XX
 PA (GENO) GEN HOSPITAL CORP.
 XX
 PI Miura M, Yuan J;
 XX
 DR WPI: 1995-051742/07.
 DR N-PSDB; AAO79969.
 XX
 PT Promoting or preventing programmed cell death in vertebrate cells
 PT - by inhibiting the activity of interleukin-1 beta converting
 PT enzyme.
 XX
 PS Claim 24; Fig 16; 116pp; English.
 XX
 CC AAO79969 encodes AAR66769 human interleukin-1 beta converting enzyme
 CC homolog (Ice-4), increasing Ice-4s enzymatic activity can
 CC promote the programmed cell death of cancer cells (pref. those
 CC overexpressing the bcl-2 oncogene), this can be used as the basis
 CC of a new cancer treatment. Alternatively by reducing Ice-4s enzymatic
 CC activity programmed cell death can be inhibited, this may be useful
 CC in the development of new cell lines which remain viable in culture
 CC for extended or indefinite periods, independent of growth factors.
 XX
 SQ Sequence 432 AA;
 Query Match 13.2%; Score 75.5; DB 16; Length 432;
 Best Local Similarity 35.6%; Pred. No. 14;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

OY 44 VFEDTVNKECYNVVDGEEL----DOEKFVVD--ENFTENYL--TDCEGK-DAGNAAGTG 94
 DB 40 VFEDLVK-----NVNGDELLKIGESASFTLNKAENLVENFLEKTDMAKIFAGHIANSQ 95
 OY 95 DE-----SDEVD 101
 DB 96 EQLSLQFSNDEDD 108

RESULT 10

AAR98464
 ID AAR98464 standard; Protein; 432 AA.

AC AAR98464;

DT 25-SEP-1996 (first entry)

DE Murine Ich-3.

XX Ich-3; murine ICE-ced-3 homologue; programmed cell death;

KW apoptosis; interleukin-1 beta converting enzyme; gene therapy.

XX Mus sp.

PN W09620721-A1.

PD 11-JUL-1996.

PF 04-JAN-1996; 96MO-US00177.

PR 04-JAN-1995; 95US-0368704.

XX (GEHO) GEN HOSPITAL CORP.

PI Miura M, Yuan J;

DR WPI: 1996-333763/33;

DR N-PSDB; AAT31554.

PT Preventing or promoting programmed cell death in vertebrate cells

PT comprises inhibiting or increasing the activity of

PT interleukin-1-beta converting enzyme, or altering expression of

PT other related genes

PS Claim 24; Fig 14; 127pp; English.

XX Ich-3 (AAR98464) causes programmed cell death and shows significant

XX homology to mouse interleukin-1 beta converting enzyme (ICE).

XX mouse mitch-2 (AAR98461) and human Ich-1 (AAR98462-63). Its sequence

XX was deduced from the Ich-3 gene (AAT31554) isolated from a mouse thymus

XX cDNA library. The protein can be obtained from host cells contg.

XX vectors that include an Ich-3 coding sequence. It can be used to

XX control the programmed cell death of vertebrate cells, to develop cell

XX lines that remain viable for extended periods, and to increase the

XX activity of ICE.

XX Sequence 432 AA;

Query Match 13.2%; Score 75.5; DB 17; Length 432;

Best Local Similarity 35.6%; Pred. No. 14;

Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

OY 44 VFEDTVNKECYNVVDGEEL----DOEKFVVD--ENFTENYL--TDCEGK-DAGNAAGTG 94

DB 40 VFEDLVK-----NVNGDELLKIGESASFTLNKAENLVENFLEKTDMAKIFAGHIANSQ 95

OY 95 DE-----SDEVD 101

DB 96 EQLSLQFSNDEDD 108

RESULT 11

ABB66916

ID ABB66916 standard; Protein; 1154 AA.

XX ABB66916;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 27540.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX W0200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001MO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

DR N-PSDB; ABL11019.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Disclosure; SEQ ID NO 27540; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA

XX sequences (AB101840-AB116175) and the encoded proteins

XX (AB57737-AB572072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1154 AA;

Query Match 12.8%; Score 73.5; DB 22; Length 1154;

Best Local Similarity 28.1%; Pred. No. 73;

Matches 27; Conservative 14; Mismatches 42; Indels 13; Gaps 4;

OY 12 NRKYDPKSPKSSDDDECK-KTGFTEYCYIVFEDTVNKECYNVVDGEELDQKFFV 69

DB 505 SKRDIIDFSKFSMDKDKLFRVAELMVRGCFLEKRMQDLGNKVCNVEEKDNVNO 564

OY 70 VDENTENYL--TDCEGKDA-----GNAAGTG 94

DB 565 LVENFWRHYIQCEGCKADKALSNPSMLGNTEYRG 600

RESULT 12

AA617789

ID AA617789 standard; Protein; 175 AA.

AC AA617789;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 18943.

XX Protein identification; signal transduction pathway; metabolic pathway;

hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
Arabidopsis thaliana.
EP1033405-A2.
06-SEP-2000.
25-FEB-2000; 2000EP-0301439.
25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126284.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
23-APR-1999; 99US-0130891.
30-APR-1999; 99US-0132048.
30-APR-1999; 99US-0132407.
04-MAY-1999; 99US-0132484.
05-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134286.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
18-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0138094.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139459.
17-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139889.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147036.
04-AUG-1999; 99US-0147204.
05-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
06-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
09-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148684.
16-AUG-1999; 99US-0148368.
17-AUG-1999; 99US-0149175.
18-AUG-1999; 99US-0149426.
20-AUG-1999; 99US-0149722.
20-AUG-1999; 99US-0149723.
20-AUG-1999; 99US-0149929.
23-AUG-1999; 99US-0149902.
23-AUG-1999; 99US-0149930.
25-AUG-1999; 99US-0150566.
26-AUG-1999; 99US-0150884.
27-AUG-1999; 99US-0151065.
27-AUG-1999; 99US-0151066.
27-AUG-1999; 99US-0151080.
30-AUG-1999; 99US-0151303.
31-AUG-1999; 99US-0151438.
01-SEP-1999; 99US-0151930.

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.5%; Score 71.5; DB 21; Length 175;
Best Local Similarity 28.6%; Pred. No. 12;
Matches 26; Conservative 12; Mismatches 32; Indels 21; Gaps 4;

OY 23 KNSDDDECKKTCFKTEYCYIVFEDTVNKECYVNVVGEEL-DOEKFYVDENFTENYLT- 80
DB 93 KKGSS-----KRALPESEFTWETDAQHK-----DAGEIHDEAVADIKEDLMSNPPLY 141
OY 81 -----DCEKGAGNAAGTGDSEDEVED 103
DB 142 FNNDADEDEFGDDGDGEDEDEDEDEED 172

RESULT 13
AAG17788
ID AAG17788 standard; Protein; 256 AA.

XX AAG17788;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 18942.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127452.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PD 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOFF/) HOFFMAN S.
XX (CARU/) CARUCCI D.
XX (GARD/) GARDNER M.
XX (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI
DR WPI: 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure: Page 283-284; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the plasmid encoded by it will help to expand
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAT70078 to AAT70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 320 AA;
SQ
Query Match 12.5%; Score 71.5; DB 21; Length 320;
Best Local Similarity 27.8%; Pred. No. 25;
Matches 27; Conservative 9; Mismatches 24; Indels 37; Gaps 5;
OY 9 FYANRKY-----TDFDKS-----FKSSDLECKK-----TCFTEYC----- 41
DB 51 FYALKKKESLLEKRRDFTSKNSNDKISIKSYDDFNEQLIITIDIKENYCLNCEGITNY 110
OY 42 ---YIVFEDIVNKECYNNVDSGELEDOEKVVDENFT 75
DB 111 DEVYITIEYENDSIL-----KDFEYFVLDKNFT 140

RESULT 15
AAB95086
ID AAB95086 standard; Protein; 589 AA.
XX
XX AAB95086;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:16999.
DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
PN

XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 16999; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 589 AA;
SQ
Query Match 12.4%; Score 71; DB 22; Length 589;
Best Local Similarity 23.8%; Pred. No. 59;
Matches 30; Conservative 23; Mismatches 37; Indels 36; Gaps 6;
OY 12 NRKYTDPEKSRFS-----SDLECKKT-----CFKTE---XCYIYFE 46
DB 158 SKFEYQNNKKEKKNIVQHTTDSLEKORTLDGSEIVKPRLECKSTRTREMSVYOLI 217
OY 47 DTNKECYNNVDSGELEDOEKVVD-----ENFTENYLTDCGKDAAGAGTGDG 97
DB 218 MTRDSGXYENSTDGEMCKDALDEDSSEVSFISGSESENETTSV-GRASDDDGSEDD 276
OY 98 DEVDED 103
DB 277 EE-DDD 281

Search completed: May 5, 2003, 17:50:37
Job time : 59 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2003, 17:45:22 ; Search time 22 Seconds

(without alignments)
137.753 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573
Sequence: 1 EBRDCCWTFYANRKYTFDK.....GKDNAGTGDSEDEVED 103

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing:

Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents:AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	80.5	14.0	126	1 US-08-044-547-3	Sequence 3, Appl
2	80.5	14.0	147	1 US-08-044-547-1	Sequence 1, Appl
3	80.5	14.0	412	2 US-08-741-134-2	Sequence 2, Appl
4	75.5	13.2	172	4 US-09-187-789-16	Sequence 16, Appl
5	75.5	13.2	172	4 US-09-139-600-11	Sequence 11, Appl
6	75.5	13.2	418	3 US-08-258-287B-58	Sequence 58, Appl
7	75.5	13.2	418	3 US-08-368-704C-56	Sequence 56, Appl
8	75.5	13.2	419	3 US-08-258-287B-57	Sequence 57, Appl
9	75.5	13.2	419	3 US-08-368-704C-55	Sequence 55, Appl
10	69.5	12.1	124	1 US-08-170-360-1	Sequence 1, Appl
11	69.5	12.1	674	4 US-08-961-083-200	Sequence 200, App
12	67	11.7	124	1 US-08-170-360-2	Sequence 2, Appl
13	67	11.7	124	2 US-08-888-497-41	Sequence 41, Appl
14	67	11.7	124	4 US-09-362-230-41	Sequence 41, Appl
15	67	11.7	124	5 PCT-US94-07926-41	Sequence 41, Appl
16	67	11.7	146	2 US-08-888-497-34	Sequence 34, Appl
17	67	11.7	146	4 US-09-362-230-34	Sequence 34, Appl
18	67	11.7	146	5 PCT-US94-07926-34	Sequence 34, Appl
19	67	11.7	299	4 US-09-069-023-20	Sequence 20, Appl
20	66.5	11.6	86	4 US-09-134-001C-3986	Sequence 3986, Ap
21	66.5	11.6	2182	2 US-08-487-826B-16	Sequence 16, Appl
22	65.5	11.4	725	2 US-08-813-940-25	Sequence 25, Appl
23	64.5	11.3	126	1 US-08-170-360-3	Sequence 3, Appl
24	64.5	11.3	126	2 US-08-888-497-38	Sequence 38, Appl
25	64.5	11.3	126	4 US-09-362-230-38	Sequence 38, Appl
26	64.5	11.3	126	5 PCT-US94-07926-38	Sequence 38, Appl
27	64.5	11.3	148	2 US-08-888-497-36	Sequence 36, Appl

28	64.5	11.3	148	4 US-09-362-230-36	Sequence 36, Appl
29	64.5	11.3	148	5 PCT-US94-07926-36	Sequence 36, Appl
30	64	11.2	519	4 US-09-172-841-55	Sequence 5, Appl
31	63	11.0	294	4 US-09-355-160D-8	Sequence 8, Appl
32	63	11.0	1658	2 US-08-609-049A-13	Sequence 13, Appl
33	63	11.0	1658	4 US-09-170-996-13	Sequence 13, Appl
34	63	11.0	1726	2 US-08-609-049A-30	Sequence 30, Appl
35	63	11.0	1726	4 US-09-170-996-30	Sequence 30, Appl
36	62.5	10.9	250	2 US-08-861-269-5	Sequence 5, Appl
37	62.5	10.9	250	3 US-09-134-596-5	Sequence 5, Appl
38	62.5	10.9	250	3 US-09-293-273-5	Sequence 5, Appl
39	62.5	10.9	1172	1 US-08-313-288B-19	Sequence 19, Appl
40	62.5	10.9	2555	3 US-08-058-489-36	Sequence 36, Appl
41	61.5	10.7	346	2 US-08-702-153-2	Sequence 2, Appl
42	61.5	10.7	346	2 US-08-702-153-4	Sequence 4, Appl
43	61.5	10.7	416	3 US-09-211-930-3	Sequence 3, Appl
44	61.5	10.7	416	3 US-09-340-993-3	Sequence 3, Appl
45	61.5	10.7	416	4 US-09-468-442-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-044-547-3
Sequence 3, Application US/08044547
Patent No. 5324715
GENERAL INFORMATION:
APPLICANT: Connolly, Thomas M.
TITLE OF INVENTION: Protein for Inhibiting
TITLE OF INVENTION: Collagen-Stimulated Platelet Aggregation
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/044,547
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/594,917
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4958
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-044-547-3

Query Match 14.0%; Score 80.5; DB 1; Length 126;
Best Local Similarity 32.7%; Pred. No. 0.21;
Matches 18; Conservative 10; Mismatches 24; Indels 3; Gaps 1;

QY 5 DCMWFYANKRYTDFDKSFKSSDLDECKKTCFKTEY---CYIVEDTVNKECYN 56
||| : : : ||| : : : ||| : : : ||| : : :
Db 45 DCMWKRPGMKLPDNLTKTEFTSVDECKKMCESAVPSCYILQINTETNECYN 99

RESULT 2
US-08-044-547-1

Sequence 1, Application US/08044547
Patent No. 5324715
GENERAL INFORMATION:
APPLICANT: Connolly, Thomas M.
APPLICANT: Keller, Paul M.
TITLE OF INVENTION: Protein for inhibiting
TITLE OF INVENTION: Collagen-stimulated Platelet Aggregation
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/044,547
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/594,917
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4958
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 147 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-044-547-1

Query Match 14.0%; Score 80.5; DB 1; Length 147;
Best Local Similarity 32.7%; Pred. No. 0.25; Mismatches 24; Indels 3; Gaps 1;

QY 5 DCMWFYANKRYTDFDKSFKSSDLDECKKTCFKTEY---CYIVEDTVNKECYN 56
||| : : : ||| : : : ||| : : : ||| : : :
Db 66 DCMWKRPGMKLPDNLTKTEFTSVDECKKMCESAVPSCYILQINTETNECYN 120

RESULT 3
US-08-741-134-2

Sequence 2, Application US/08741134
Patent No. 5861498
GENERAL INFORMATION:
APPLICANT: Litwack, Gerald
APPLICANT: Alnemri, Emad S.
APPLICANT: Fernandez-Alnemri, Teresa
TITLE OF INVENTION: IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
TITLE OF INVENTION: AND
METHODS OF USING THE SAME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5861498ris
STREET: One Liberty Place - 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 3.11
SOFTWARE: Wordperfect for Windows 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/741,134
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,163
FILING DATE: 01-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-2090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-741-134-2

Query Match 14.0%; Score 80.5; DB 2; Length 412;
Best Local Similarity 30.6%; Pred. No. 0.87; Mismatches 29; Indels 27; Gaps 5;

QY 11 ANRKYTDFDKSFKSS---SDLECKKTCFKTEYCYIVEDTVNKECYNVVDGEELD 65
||| : : : ||| : : : ||| : : : ||| : : :
Db 133 ANRKAQPKDKAKGNKSAFAESDSDDD-----EDQLK-----FLDGEIDIT 174

QY 66 KPVVDENFTENTYLTDCGKDAGNAGTGDSEVEDD 103
||| : : : ||| : : : ||| : : : ||| : : :
Db 175 DE--NDESFKN--TSAEGDSDSEDDDEDEDEDD 208

RESULT 4
US-09-187-789-16

Sequence 16, Application US/09187789
Patent No. 6340740
GENERAL INFORMATION:
APPLICANT: Alnemri, Emad S.
APPLICANT: Fernandez-Alnemri, Teresa
TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
TITLE OF INVENTION: AND METHODS OF USE
FILE REFERENCE: 480140.434C1
CURRENT APPLICATION NUMBER: US/09/187,789
CURRENT FILING DATE: 1998-11-06
NUMBER OF SEQ ID NOS: 78
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 16
LENGTH: 172
TYPE: PRT
ORGANISM: Mus musculus
US-09-187-789-16

Query Match 13.2%; Score 75.5; DB 4; Length 172;
Best Local Similarity 35.6%; Pred. No. 1; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDTVNKECYNVVDGEEL---DQKFFVD--ENTETNYL--TDEGK--DAGNAGTG 94
||| : : : ||| : : : ||| : : : ||| : : :
||| : : : ||| : : : ||| : : : ||| : : :

RESULT 9
US-08-368-704C-55

STATE: D. C.

COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/170,360
FILING DATE: 03-MAR-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00333
FILING DATE: 06-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 7058
FILING DATE: 04-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Ernst, Barbara G.
REGISTRATION NUMBER: 30,377
REFERENCE/DOCKET NUMBER: 1871-104A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-170-360-1

Query Match 12.1%; Score 69.5; DB 1; Length 124;
Best Local Similarity 27.1%; Pred. No. 2.9;
Matches 16; Conservative 8; Mismatches 20; Indels 15; Gaps 1;

QY 28 LDECKKTCFTEYCYIYFEDTVNKECYNNVDSGELDQEKYVDENTENYVLDCEKD 86
Db 38 VDELDRCETHDNCY-----RDANKLDSCFKLVNDPYTESYSNSNTE 81

RESULT 11
US-08-961-083-200
Sequence 200, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 200:
SEQUENCE CHARACTERISTICS:
LENGTH: 674 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-200

Query Match 12.1%; Score 69.5; DB 4; Length 674;
Best Local Similarity 23.6%; Pred. No. 23;
Matches 25; Conservative 17; Mismatches 35; Indels 29; Gaps 5;

QY 7 WTYANRKYTDPEKSKSDLDCEKT-----CFTEYCYIYFEDTVNKECYNNVDSGE 61
Db 322 WYQNKRLKEFD-----SEIINCRTSEQISCYSDISYTF-----LRYFADPV 368

QY 62 ELDOEKF-----VDENFTENYVLDCEG-----KDAGNAAGTGE 96
Db 369 QEDKALYDODLVYTKNDDLPATDLDYPLAAYRFGGRATFGQE 414

RESULT 12
US-08-170-360-2
Sequence 2, Application US/08170360
Patent No. 5656602
GENERAL INFORMATION:
APPLICANT: Tseng, Albert P. S.
APPLICANT: Ingllis, Adam
TITLE OF INVENTION: PLAZ INHIBITORY COMPOUNDS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg Ernst & Kurz
STREET: Suite 701-E, 535 Thirteenth St., N.W
CITY: Washington
STATE: D. C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/170,360
FILING DATE: 03-MAR-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00333
FILING DATE: 06-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 7058
FILING DATE: 04-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Ernst, Barbara G.
REGISTRATION NUMBER: 30,377
REFERENCE/DOCKET NUMBER: 1871-104A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein


```

; Sequence 41, Application PC/TUS9407926
; GENERAL INFORMATION:
; APPLICANT: Tischfield, Jay A.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide
; TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences
; TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide
; TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
; ADDRESSEE: Russell PA
; STREET: 200 East Broward Boulevard
; CITY: Fort Lauderdale
; STATE: FL
; COUNTRY: USA
; ZIP: 33301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07926
; FILING DATE: 13-JUL-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/097,354
; FILING DATE: 26-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Manso, Peter J.
; REGISTRATION NUMBER: 32,264
; REFERENCE/DOCKET NUMBER: IN21044-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 305-527-2498
; TELEFAX: 305-764-4996
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-07926-41

```

Query Match 11.7%; Score 67; DB 5; Length 124;

Best local Similarity 22.7%; Pred. No. 5.4;

Matches 17; Conservative 14; Mismatches 22; Indels 22; Gaps 3;

```

QY 27 DLDECKKTCRTEYCYIVEDFVNKECYNVVDGEELDQEKFYVDENFTENTYLTDCGKRD 86
   ||| | | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 40 DLDRCC--CQTHDHCY-----NQAKKLECKFLIDNPYNTYSYKC----- 77
QY 87 AGNAAGTGESDEVD 101
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 78 SGNVITCSDKNDCE 92

```

Search completed: May 5, 2003, 17:51:12
Job time : 23 secs


```

Query Match:                               13.28; Score 75.5; DB 9; Length 172;
Best Local Similarity 35.68; Pred No.1.5; Mismatches 14; Indels 19; Gaps 6
Matches    26; Conservative    14;
QY      44 VFEDVYNKECYNVVDGEEL---DOEKRYVD--ENFTENTYL--TDCEGK-DAGNNAATG 94
        |||.|||   |||.||| | :|:| | |||.||| | | | | |
Db      26 VFDDIVLEK-----NLNGDELLKIGESASPIILUKAENVLENTLEKTDMAGRIFAGHANS 81
QY      95 DE-----SDEVYD 101
        :::::| || |
Db      82 EQSLQFSNDEDD 94

RESULT 2
US-09-989-903-16
; Sequence 16, Application US/09989903
; Patent No. US20020146804A1
; GENERAL INFORMATION:
; APPLICANT: Alnemati, Emad S.
; APPLICANT: Fernandez-Alnemati, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APTOTIC PROTEASE, NUCLEIC ACID ENCODING
; TITLE OF INVENTION: AND METHODS OF USE
```

```
FILE REFERENCE: 480140.434D1
CURRENT APPLICATION NUMBER: US/09/989,903
CURRENT FILING DATE: 2002-04-11
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 16
LENGTH: 172
TYPE: PRF
ORGANISM: Mus musculus
US-09-989-903-16

Query Match
Best Local Similarity 35.6%; Score 75.5; DB 10; Length 172;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDTVNKECYVNVDEEL---DQKRYVD--ENTENTL--TDCEK-DAGNAGTG 94
DB 26 VFDDIVER-----NVLNGDELKIGESASFLINKAENVLEKTMDWAGIFAGHANSQ 81
QY 95 DE-----SDEVD 101
DB 82 EQLSLQFSNDEDD 94

RESULT 3
US-09-833-790-252
Sequence 252, Application US/09833790
GENERAL INFORMATION:
APPLICANT: Lodes, Michael J.
APPLICANT: Wang, Tonglong
APPLICANT: Secrist, Heather
APPLICANT: Mohamath, Raodoh
APPLICANT: Indrias, Carol Y.
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.512
CURRENT APPLICATION NUMBER: US/09/833,790
CURRENT FILING DATE: 2001-04-11
NUMBER OF SEQ ID NOS: 440
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 252
LENGTH: 156
TYPE: PRF
ORGANISM: Homo sapiens
US-09-833-790-252

Query Match
Best Local Similarity 12.3%; Score 70.5; DB 10; Length 156;
Matches 26; Conservative 20; Mismatches 50; Indels 23; Gaps 6;

QY 1 EERE-----DCWTFYANKRYTDFDKS---FKSSD-----LDECKKTFKRYCYIV 44
DB 30 EEOEKLLKRSCTLYGNLSFYTTEQIYELPSKSGDIKRIIMGDKMKKTA--CGCFEVE 87
QY 45 FEDTVNKECYVNVDEELDQKRYVDENTENTLTDCEK-DAGNAGTGESDEVED 103
DB 88 YTSRADAEAMRYINGTRLDLR--IIRIDWAGFR---EGQYGRGSGGVREYRKPD 141

RESULT 4
US-09-815-242-11522
Sequence 11522, Application US/09815242
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
```

```
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11522
LENGTH: 1167
TYPE: PRF
ORGANISM: Helicobacter pylori
US-09-815-242-11522

Query Match
Best Local Similarity 12.2%; Score 70; DB 10; Length 1167;
Matches 24; Conservative 12; Mismatches 24; Indels 18; Gaps 4;

QY 19 DKFSKSSDDECKKTCFTECYIVFEDTVNKECYVNVDEELDQKRYVDENTENTY 78
DB 240 DKTEQSSNDPKENQKTLFDA---ILLQDLAN--AYVNMPTK-----LDDRWWEWF 286
QY 79 LTDCEK-DAGNAGTGDE 96
DB 287 -----AKTGNIRARTLINE 299

RESULT 5
US-09-765-272-200
Sequence 200, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
```



```

      PRIOR FILING DATE: 2000-10-11
      NUMBER OF SEQ ID NOS: 10
      SOFTWARE: PatentIn version 2.1
      SEQ ID NO 5
      LENGTH: 148
      TYPE: PRT
      ORGANISM: Homo sapiens
      US-09-975-456B-5

Query Match                               11.3%; Score 64.5; DB 9; Length 148;
Best Local Similarity    29.8%; Pred. No. 18;
Matches    17; Conservative     5; Mismatches    20; Indels    15; Gaps    2;

QY      28 LDDEKTKCTECFYVFEEDTVNKKCCYYNVVDGEELDDEKFVNDFENTFYLDCDG 84
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       68 VDDELKCCQTHNCY-----DQAK-----LDSCFKLLDNPTHTTYSYCSG 101

RESULT 12
US-09-925-297-511
Sequence 511, Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925,297
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 511
LENGTH: 156
TYPE: PRT
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: SITE
LOCATION: (156)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-925-297-511

Query Match                               11.3%; Score 64.5; DB 10; Length 156;
Best Local Similarity    29.8%; Pred. No. 19;
Matches    17; Conservative     5; Mismatches    20; Indels    15; Gaps    2;

QY      28 LDDEKTKCTECFYVFEEDTVNKKCCYYNVVDGEELDDEKFVNDFENTFYLDCDG 84
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       68 VDDELKCCQTHNCY-----DQAK-----LDSCFKLLDNPTHTTYSYCSG 109

RESULT 13
US-10-102-806-695
Sequence 695, Application US/10102806
Publication No. US20030054421A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA103P1C1
CURRENT APPLICATION NUMBER: US/10/102,806
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: 09/925,298
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05881
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 846
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 695
LENGTH: 460

```

Tue May 6 16:59:48 2003

us-09-936-737a-2.rapb

Page 5

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-102-806-695

```

Query Match	11.28;	Score 64;	DB 9;	Length 460;
Best Local Similarity	25.88;	Pred. No. 72;		
Matches` 34;	Conservative 21;	Mismatches 45;	Indels 32;	Gaps 77;

QY 3 REDCWTFYANR--KYTDFDKSFKKSSDL--DECKKTCFKEICYIVFEDTVNKECYINV 57

DD 298 REDDLILFSNHGELKWLDFVGAKEG.LLFKEKAKEALGAKKDANNGNLQLRNKEVIMEV 35/

[illegible]

Db 358 LEG-EVEKEALKKIIDQESLNKWSKSGRRPKGKGKGNKAQPGSGKGKVFQGGKTKF 416

```

95  --DEVELOPED 103
0Y  |||||:

```

Db 417 ASDDEHDEHDEH 428

RESULT 14
US-10-042-417-52

; Sequence 52, Application US/10042417
; Patent No. US20020123082A1
; GENERAL INFORMATION.

; GENERAL INFORMATION:
; APPLICANT: Pagano, I

```

; TITLE OF INVENTION: METHODS TO IDENTIFY COMPOUNDS USEFUL FOR THE TREATMENT OF
; TITLE OF INVENTION: PROLIFERATIVE AND DIFFERENTIATIVE DISORDERS
;
;
;

```

CURRENT APPLICATION NUMBER: US/10/042,417
CURRENT FILING DATE: 2003-01-07

PRIOR APPLICATION NUMBER: 60/260,179

;; PRIOR FILING DATE: 2001-01-5
; NUMBER OF SEQ ID NOS: 89

; SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO 52
;
; LENGTH: 590

```

```
TYPE: PRT
```

ORGANISM: Homo sapiens

NAME/KEY: SITE

LOCATION: all Xaa posi

OTHER INFORMATION: Xaa
US-10-042-417-52

100

Query Match	Post Local Similarity
1	0.95
2	0.85
3	0.75
4	0.65
5	0.55
6	0.45
7	0.35
8	0.25
9	0.15
10	0.05

Best Local Similarity
Matches 23; Conservat

100

```

qy      16 TDFDKSFKKSSDLDEC
          | | | |
pb      356 TSHOCCETKTS-TSKT

```

[illegible]

RESULT 15
US-10-092-219-8

; Sequence 8, Application US/10092219
; Patent No. US20020115114A1
; GENERAL INFORMATION

APPLICANT: DOMIN, Jan

TITLE OF INVENTION: No. US20020115114A1e1 Lipid Kinase

FILE REFERENCE: 1064HG/50947
CURRENT APPLICATION NUMBER: HS/10/093 319

CONTENT REFLECTION NUMBER: 03/10/032,213
CURRENT FILING DATE: 2002-03-07

PRIOR APPLICATION NUMBER: PCT/GB98/00244

PRIOR FILING DATE: 1998-01-27
PRIOR APPLICATION NUMBER: 09/355.160

PRIOR FILING DATE: 1999-10-01

```

; PRIOR APPLICATION NUMBER: 9701652.1
; PRIOR FILING DATE: 1997-01-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8

```

ORGANISM: Homo sapiens
US-10-092-219-8

Query Match	11.08;	Score 63;	DB 12;	Length 294;
Best Local Similarity	38.89;	Prod No	55.	

Best Local Similarity	28.88	Pred. NO. 33							
Matches	19	Conservative	18	Mismatches	25	Indels	4	Gaps	2

QY 12 NRRYTFDK

```

:|:|:
45 DROMTSPB
Dh

```

Search completed: May 5, 2003, 17:54:15
Job time : 36 secs

QY 67 KEVDENFTENYLTDCGKADGNAAGTGDSEDEVD 102
 Db 428 KDEKEGDEGDGDDGDEDESE--DEGDEGDEGE 461

RESULT 7

T22054
 hypothetical protein F40G12.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2000
 C:Accession: T22054
 R:McMurray, A.
 submitted to the EMBL Data Library, July 1996
 A:Reference number: Z19507
 A:Accession: T22054
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-264 <WIL>
 A:Cross-references: EMBL:Z77661; PIDN:CA801183.1; GSPDB:GN00023; CESP:F40G12.5
 A:Experimental source: clone F40G12
 C:Genetics:
 A:Gene: CESP:F40G12.5
 A:Map position: 5
 A:Introns: 57/3; 158/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein F40G12.4

Query Match 13.2%; Score 75.5; DB 2; Length 264;
 Best Local Similarity 27.7%; Pred. No. 8.8;
 Matches 31; Conservative 19; Mismatches 33; Indels 29; Gaps 8;

QY 1 EEREDCWTFYANKRYTD-----FDK-SFKSSDLDECKKTCFTECYIVFEDYVNVK 52
 Db 29 ESESDC-----FEKVFLLISGKHECSK-----DYDFLRNLIQRREALTSGKE 72

QY 53 CYYNVVDGEELDOEKF-VVDENFTE--NYLTDCGKADGNAAGTGDSEDEVD 101
 Db 73 CLELEIV-EECPKEKFLIENYSQLVLLTE-KPRNGACTAPVFLQLEIE 122

RESULT 8

T16354
 hypothetical protein F42G9.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16354
 R:Raich, A.
 submitted to the EMBL Data Library, March 1996
 A:Description: The sequence of C. elegans cosmid F42G9.
 A:Reference number: Z18498
 A:Accession: T16354
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-491 <TAI>
 A:Cross-references: EMBL:U00051; NID:G1216305; PID:G1216310; PIDN:AAA91358.1; CESP:F42G9
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F42G9.1
 A:Introns: 37/3; 254/2; 364/2; 406/2

Query Match 13.2%; Score 75.5; DB 2; Length 491;
 Best Local Similarity 24.5%; Pred. No. 16;
 Matches 23; Conservative 18; Mismatches 28; Indels 25; Gaps 4;

QY 10 YANRKYTF--DKSFKSSDLDECKKTCFTECYIVFEDYVNVKCYYNVVDGEELDOEK 67
 Db 67 FTSAKLPFLKRFKFWEDADVAECLQKAF-----VDFDFIRAE-----ESMKELK 112

QY 68 FVVDENFTENYLTDCGKADGNAAGTGDSEDEVD 101
 Db 113 DIGDEG-----KKPKAGGADSDSEADRID 137

RESULT 9

T34510
 hypothetical protein ZK1290.10 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T34510
 R:Raich, A.
 submitted to the EMBL Data Library, July 1995
 A:Description: The sequence of C. elegans cosmid ZK1290.
 A:Reference number: Z21535
 A:Accession: T34510
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-360 <TAI>
 A:Cross-references: EMBL:U21308; PIDN:AAB93316.1; GSPDB:GN00020; CESP:ZK1290.10
 A:Experimental source: strain Bristol N2; clone ZK1290
 C:Genetics:
 A:Gene: CESP:ZK1290.10
 A:Map position: 2
 A:Introns: 23/1; 48/1; 158/3; 216/1; 264/1; 312/1; 331/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein ZK1290.10

Query Match 13.1%; Score 75; DB 2; Length 360;
 Best Local Similarity 33.3%; Pred. No. 13;
 Matches 23; Conservative 9; Mismatches 23; Indels 14; Gaps 3;

QY 46 EDYVNECYYNVVDGEELDOEKFYV-----DENFTENYLTDCGKADGNAAGT 94
 Db 224 EDYVGE--NVDPEKFLFKIYIRNDGSTTENTQSTTEKTKTIEVSVELGEG 280

QY 95 DESDEVED 103
 Db 281 DDEDEND 289

RESULT 10

T18402
 aspartate/aspartate rich protein - malaria parasite (Plasmodium falciparum) (fragment)
 C:Species: Plasmodium falciparum
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18402
 R:Barale, J.C.; Candèlle, D.; Attal-Bonnefoy, G.; Dehoux, P.; Bonnefoy, S.; Ridley, R.
 Infect. Immun. 65, 3003-3010, 1997
 A:Title: Plasmodium falciparum AAP1, a giant protein containing repeated motifs rich
 A:Reference number: Z18929; MID:97378065; PMID:9234746
 A:Accession: T18402
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3844 <BAR>
 A:Cross-references: EMBL:Y08926; NID:e1154302; PID:e311435; PIDN:CAA70130.1
 C:Genetics:
 A:Gene: aarp1

Query Match 13.1%; Score 75; DB 2; Length 3844;
 Best Local Similarity 23.9%; Pred. No. 1.5e+02;
 Matches 21; Conservative 23; Mismatches 34; Indels 10; Gaps 3;

QY 1 EEREDCWTFYANKRYTD-----FDK-SFKSSDLDECKKTCFTECYIVFEDYVNVK 51
 Db 1726 EDSEECQIKSKYKMSKCNKNENIIFDSISVLKRNKIKRLKYMCKKNKCIYIDDNNK 1785

QY 52 EGYYNVVDGEELDOEKFYVDENFTENYL 79
 Db 1786 K-KNNKKNVNDQEKRYVLNKKIFVHNFI 1812

RESULT 11

T32731
 PAR interacting protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T32731
 R:Comte, P.A.; Ossipow, V.; Schibler, U.
 submitted to the EMBL Data Library, January 1997

A:Description: Isolation of PIP, a 160 kDa nucleolar protein that interacts with the act
A:Reference number: Z21213
A:Accession: T32731
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1277 <CON>
A:Cross-references: EMBL:083590; NID:g2253210; PID:g2253211; PIDN:AA62878.1

Query Match 13.0%; Score 74.5; DB 2; Length 1277;
Best Local Similarity 30.4%; Pred. No. 53;
Matches 31; Conservative 9; Mismatches 37; Indels 25; Gaps 5;

QY 1 EEREDCWFYANKRYDFD-KSFKSSDDECKKTCFTEYCYVFEEDYVNEKYNNVD 59
DB 630 EDEED-----NYYVFTDDEKOLKHGEDADS-----EDSKSESDVSD 670
DB 671 GESEEDRDKVDPRQOLM---EVLQGNALGGESEEE 709

RESULT 12

D70465
hypothetical protein aq_1923 - Aquifex aeolicus

C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: D70465
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oy
V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: D70465

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-296 <AG>

A:Cross-references: GB:AE000763; NID:g2984178; PIDN:AAC07719.1; PID:g2984187; GB:AE00065

A:Experimental source: strain VFS

C:Genetics:

A:Gene: aq_1923

C:Superfamily: Aquifex aeolicus hypothetical protein aq_1923

Query Match 12.9%; Score 74; DB 2; Length 296;
Best Local Similarity 24.2%; Pred. No. 14;
Matches 22; Conservative 18; Mismatches 21; Indels 30; Gaps 4;

QY 9 FYANKRYD-----FDKSKSSDDECKKTCFTEYCYVE---DYNKCY 54
DB 18 FYADKYDNKAKAQVEFLDKTLKSG-----KSKYVYDKPIGGLIHKDV 67
QY 55 YNVVDGEELDOEKFEVDENFTENYLTDOEGK 85
DB 68 YRDRNGEEFKIEELIEK-----LSFECK 92

RESULT 13

B70236

hypothetical protein BBH09 - Lyme disease spirochete plasmid H/Jp28-3

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999

C:Accession: B70236

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:98065943; PMID:9403685

C:Accession: B70236

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1278 <RLF>

A:Cross-references: GB:AE000784; NID:g2690041; PIDN:AAC66000.1; PID:g2690056; TIGR:BBH09

A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid

Query Match 12.9%; Score 74; DB 2; Length 1278;
Best Local Similarity 25.0%; Pred. No. 59;
Matches 18; Conservative 17; Mismatches 25; Indels 12; Gaps 3;

QY 10 YANRRYTFDKSKFKSSDDECKKTCFTEYCYVFEEDYVNEKYNNV-----VDGEELD 64
DB 181 YEKRYIED--FSKIEKE-----YKEGFILSYLIRKRYLKTSLIYSKEQIS 233
QY 65 QEFYVDENFTE 76
DB 234 KEKEIIOKTLRE 245

RESULT 14

S17231
dynein beta heavy chain, ciliary - sea urchin (Anthocidaris crassispina)

N:Contains: dynein ATPase (EC 3.6.4.2)

C:Species: Anthocidaris crassispina

C:Date: 30-Sep-1991 #sequence_revision 02-May-1994 #text_change 19-Apr-2002

C:Accession: S17231; PS0415

R:Ogawa, K.

Nature 352, 643-645, 1991

A:Title: Four ATP-binding sites in the midregion of the beta heavy chain of dynein.

A:Reference number: S17231; MUID:91326104; PMID:1830928

A:Accession: S17231

A:Molecule type: mRNA

A:Residues: 1-4466 <OGA>

A:Cross-references: GB:D01021; NID:g217202; PIDN:BAA0827.1; PID:g217203

R:Ogawa, K.

Proc. Jpn. Acad. B Phys. Biol. Sci. 67, 27-31, 1991

A:Title: ATP-binding site in dynein beta-heavy chain: identification by molecular clo

A:Reference number: PS0415

A:Accession: PS0415

A:Molecule type: mRNA

A:Residues: 764-1001, 'APQ', 1005-2036, 'VPSSVET' <OG2>

C:Superfamily: dynein heavy chain, ciliary

C:Keywords: ATP; heterotetramer; hydrolase; microtubule binding; nucleotide binding;

F:154-161/Region: nucleotide-binding motif A (P-loop)

F:1852-1859/Region: nucleotide-binding motif A (P-loop)

F:2133-2140/Region: nucleotide-binding motif A (P-loop)

F:2460-2467/Region: nucleotide-binding motif A (P-loop)

F:2805-2812/Region: nucleotide-binding motif A (P-loop)

F:160/Binding site: ATP (Lys) #status predicted

F:1858/Binding site: ATP (Lys) #status predicted

F:2139/Binding site: ATP (Lys) #status predicted

F:2466/Binding site: ATP (Lys) #status predicted

F:2811/Binding site: ATP (Lys) #status predicted

Query Match 12.9%; Score 74; DB 1; Length 4466;
Best Local Similarity 25.9%; Pred. No. 2,1e+02;
Matches 28; Conservative 14; Mismatches 28; Indels 38; Gaps 6;

QY 1 EEREDCWFYANKRY-----TDFKSKFKSSDIDE-----CK-----KTC 35
DB 460 EEFQECAKVTEPRYDGLPTCOEFLEDEYEFKFKFDDRLGSLICGFPDCCGLEMA 519
QY 36 FTEYCY-----YFETVNEKYNNV-----DGEELDOEKFEVDEN 73
DB 520 FKMLDCYGLDPRVIRNDF---ECKYPIVLMYDQELDQSKREIYDH 564

RESULT 15

T21428

hypothetical protein F26F2.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21428

R:Ainscough, R.

submitted to the EMBL Data Library, November 1996

Tue May 6 16:59:49 2003

us-09-936-737a-2.rpr

Page 5

A;Reference number: Z19420

A;Accession: T21428
A;Status: preliminary

A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA

A;Molecule type: DNA
A;Residues: 1-643

A;Residues: 1-643 <WIL>
A;Cross-references: EMB

A/Cross-references: EMBL:Z81514; PIDN:CAB04192.1; GSPDB:GN00023; CESP:F26F2.6
A/Experimental source: clone F26F2

A; Experimental source: clone F26F2
C: Genetics.

C; Genetics :
A: Gene : CFS

A:Gene: CESP:F26F2.6
A:Map position: 5

A:Map position: 5
A:Introns: 18/3; 116/3; 142/1; 265/3; 306/3; 373/3; 403/2; 425/2; 479/3; 544/3; 573/2; 6

Query Match 12 78: Score 73: DB 2: Tomath 643:

Query Match	12.7%;	Score 73;	DB 2;	Length 643;
Best Local Similarity	26.6%;	Pred. No. 37.		

Best Local Similarity 26.6%; Pred. No. 37;
Matches 29: Conservative 14: Mismatches

Matches 29; Conservative 14; Mismatches 32; Indels 34; Gaps 6;

QY 7 WTFY-----ANRKYT---DFDKSFKKS-----SDLDECKKTCFTEYCYIVED 47

Db 155 WSEFRPYVDGCDANCSVTMVQDYAKPGPETVKPRDAQISNWDECMHMCYEEECQIAYLD 214

QY 48 TYNKEC-YYNVGDGEELDQE-----KVVNDEN---FTENYLYTD 81

Db 215 SDAKCVWYSSDDGLTFMNKSSADSGKRLVIKMKVDETTCLEFPPQLLD 263

```
Search completed: May 5, 2003, 17:53:28
Job time : 40 secs
```


GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2003, 17:20:27 ; Search time 21 seconds

(Without alignments)
203.432 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573
Sequence: 1 EERDCWTFYANKRYTDFDK.....GKDAAGNACTGDESDDEVDD 103

Scoring table: BLASTUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	85.5	14.9	147	1 LAP_HABOF	Q01747 haementeria
2	80.5	14.0	412	1 FRB4_SPOFR	Q26486 spodioptera
3	78	13.6	797	1 VG48_HSVSA	Q01033 herpesvirus
4	75.5	13.2	419	1 ICEC_MOUSE	Q08736 mus musculus
5	75.5	13.2	491	1 P2C1_CAEEL	P49595 caenorhabdi
6	75	13.1	360	1 YOP4_CAEEL	Q09337 caenorhabdi
7	74	12.9	296	1 Y123_AQDAE	Q67757 aquilex aeo
8	73	12.9	4466	1 DHQC_ANTCR	P39057 anthocidari
9	73	12.7	1405	1 DPOA_SCHPO	P38040 schizosacch
10	71	12.4	782	1 YNV2_YEAST	P40157 saccharomyc
11	70.5	12.3	78	1 MOH1_CANGP	P81034 cancer pagu
12	70.5	12.3	146	1 PA21_CANGP	P06596 canis famil
13	70.5	12.3	156	1 CB20_HUMAN	P52288 homo sapien
14	70.5	12.3	748	1 Y875_METVA	Q58285 methanococc
15	70	12.2	1058	1 S185_YEAST	P40856 saccharomyc
16	69.5	12.1	146	1 PA21_PIG	P00592 sus scrofa
17	68.5	12.0	610	1 CALG_HUMAN	Q14967 homo sapien
18	68	11.9	146	1 PA21_MOUSE	Q92097 mus musculu
19	68	11.9	240	1 NEF_HV2D2	P15829 human immun
20	68	11.9	4466	1 DYHC_TRIGR	P23098 tritipeustes
21	67.5	11.8	277	1 PS12_ARATH	Q33712 arabidopsis
22	67.5	11.8	321	1 D7_AEDAE	P18155 aedes aegypt
23	67.5	11.8	602	1 SPT8_YEAST	P38935 saccharomyc
24	67	11.7	146	1 PA21_RAT	P04035 rattus norv
25	67	11.7	327	1 RRP_PIRIV	Q01769 piri virus
26	67	11.7	650	1 CC45_YEAST	Q08033 saccharomyc
27	67	11.7	2410	1 MOK1_SCHPO	Q08033 schizosacch
28	66.5	11.6	78	1 MOH2_CANGP	P81035 cancer pagu
29	66.5	11.6	125	1 FABL_ANOPU	Q90238 anolis pulc
30	66.5	11.6	132	1 PA21_HORSE	P00594 equus cabal
31	66.5	11.6	405	1 LHX1_BRANE	Q90476 brachydanio
32	66.5	11.6	591	1 CALX_RAT	P35565 rattus norv
33	66.5	11.6	733	1 YV08_YEAST	P46995 saccharomyc

34	66.5	11.6	1033	1 S190_YEAST	P36123 saccharomyc
35	66	11.5	148	1 CB20_XENIA	P52299 xenopus lae
36	66	11.5	1647	1 YDE1_SCHPO	Q10435 schizosacch
37	65.5	11.4	165	1 ADP1_CAEEL	Q07750 caenorhabdi
38	65.5	11.4	428	1 RF4_KILOA	P09806 kluyveromyc
39	65.5	11.4	455	1 YNU6_YEAST	P40161 saccharomyc
40	65.5	11.4	725	1 RIR1_MYCTU	P50640 mycobacteri
41	65.5	11.4	746	1 GYP7_YEAST	P48365 saccharomyc
42	65.5	11.4	793	1 NEFB_DROME	Q24568 drosophila
43	65	11.3	1124	1 TRP8_HUMAN	P37275 homo sapien
44	65	11.3	1281	1 TRPS_HUMAN	Q9uhf7 homo sapien
45	65	11.3	1616	1 VIT1_CAEEL	P5155 caenorhabdi

ALIGNMENTS

```

RESULT 1
LAP_HABOF STANDARD; PRT; 147 AA.
ID LAP_HABOF
AC Q01747;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE Anti-platelet protein precursor.
GN LAP.
OS Haementeria officinalis (Mexican leech).
OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinea; Hirudinea;
OC Rhynchobdellida; Glossiphoniidae; Haementeria.
OX NCBI_TaxID=6410;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 60-91 AND 123-139.
RC TISSUE=Salivary gland.
RX MEDLINE=92202247; Pubmed=1551898;
RA Keller P.M., Schultz L.D., Condra C., Karczewski J., Connolly T.M.;
RT "An inhibitor of collagen-stimulated platelet activation from the
RT salivary glands of the Haementeria officinalis leech. II. Cloning of
RT the cDNA and expression."
RT J. Biol. Chem. 267:6899-6904(1992).
CC -!- FUNCTION: AN INHIBITOR OF COLLAGEN-STIMULATED PLATELET
CC -!- SUBCELLULAR LOCATION: DENSE GRANULE RELEASE AND SEROTONIN RELEASE.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M81489; AAA29194.1; -.
KW Signal.
FT SIGNAL. 1 21 POTENTIAL.
FT CHAIN 22 147 ANTI-PLATELET PROTEIN.
FT SEQUENCE 147 AA; 15908 MW; 75A55113744EA2E CRC64;
SQ
Query Match 14.9%; Score 85.5; DB 1; Length 147;
Best Local Similarity 33.9%; Pred. No. 0.2;
Matches 19; Conservative 10; Mismatches 24; Indels 3; Gaps 1;
QY 4 EDCWTFYANKRYTDFDKSFKSSDLECKTCFTEF---CYIYEDTVAKKCYNN 56
Db 65 EDCWTKRPKRWKLPDNLITKTEFTSVDCRKRCKESAVEPCYIIQINTNECYRN 120
RESULT 2
FKB4_SPOFR STANDARD; PRT; 412 AA.
AC Q26486;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)

```

DE 46 kDa FK506-binding nuclear protein (Peptidyl-prolyl cis-trans
DE isomerase) (PPIase) (EC 5.2.1.8).
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Noctuidae; Noctuidae; Amphipyrinae; Spodoptera.
OX NCBI_TaxID=7108;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=95074110; PubMed=7527037;
RA Alnemri E.S., Fernandes-Alnemri T., Pomeroy K., Robertson N.M.,
RA Dudley K., Dubois G.C., Litwack G.;
RT "FKBP46, a novel sf9 insect cell nuclear immunophilin that forms a
RT protein-kinase complex";
RL J. Biol. Chem. 269:30828-30834(1994).
CC -1- FUNCTION: BINDS TO, AND IS INHIBITED BY FK506 AND RAPAMYCIN. BINDS
CC DOUBLE-STRANDED DNA IN VITRO. PPIASES ACCELERATE PROTEIN FOLDING.
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PPM: PHOSPHORYLATED BY A NUCLEAR KINASE IN THE PRESENCE OF MG2+
CC AND ATP.
CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: U15038; AAA58962.1; -
DR HSSP: P20071; 1FKJ.
DR InterPro: IPR001179; FKBP_PPIase.
DR Pfam: PF00254; FKBP; 1.
DR PROSITE: PS00453; FKBP_PPIASE_1; FALSE_NEG.
DR PROSITE: PS00454; FKBP_PPIASE_2; 1.
DR PROSITE: PS50059; FKBP_PPIASE_3; 1.
KW isomerase; Notomase; Nuclear protein; DNA-binding; Phosphorylation.
FT DOMAIN 91 112 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 120 145 LYS-RICH (BASIC).
FT DOMAIN 152 216 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 219 302 LYS-RICH (BASIC).
FT DOMAIN 324 412 PPIASE, FKBP-TYPE.
SQ SEQUENCE 412 AA; 45810 MW; F2A69159AEFE22 CRC64;
Query Match 14.0%; Score 80.5; DB 1; Length 412;
Best Local Similarity 30.6%; Pred. No. 1.7;
Matches 30; Conservative 12; Mismatches 29; Indels 27; Gaps 5;
QY 11 ANKRYDFOKSEFKS-----SDLECKKTCFTECYIVFEDTVNKECYVNVGDEELDQ 65
DB 133 ANKRAPDKKAGKNSAPAEASDSDDD-----EDQLK-----FLDESDIDT 174
QY 66 EKFEVDNFTENYLTDCGKDGAGNAAGTGDSESDVE 103
DB 175 DE--NDSFKRM--TSAEGDSDSEDDDEDEDEDD 208
RESULT 3
VG48_HSVSA STANDARD: PRT; 797 AA.
ID VG48_HSVSA
AC 001033;
DT 01-APR-1993 (rel. 25, Created)
DT 01-APR-1993 (rel. 25, Last sequence update)
DT 01-APR-1993 (rel. 25, Last annotation update)
DE Hypothetical gene 48 protein.
GN 48 OR EDLFS.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;

RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=9233688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Bisinger B.,
RA Newman C., Wilmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome";
RL J. Virol. 66:5047-5058(1992).
CC -1- SIMILARITY: TO EBV BRRE2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: X64346; CAA45671.1; -
DR PIR: A36811; A36811.
KW Hypothetical protein.
SQ SEQUENCE 797 AA; 88667 MW; 9E294234AD850E23 CRC64;
Query Match 13.6%; Score 78; DB 1; Length 797;
Best Local Similarity 32.3%; Pred. No. 5.7;
Matches 31; Conservative 12; Mismatches 33; Indels 20; Gaps 5;
QY 11 ANKRYDFOKSEFKS-----SDLECKKTCFTECYIVFEDTVNKECYN-----VVDGEELDQ 66
DB 382 ANEK--EKRIIDSDDDDDKDKDEYLE-----NE--YNRDEDEDEDEDE 427
QY 67 KFEVDNFTENYLTDCGKDGAGNAAGTGDSESDVE 102
DB 428 KDEKEGEDEGDGDEGEDEGE--DEGDEDEGE 461
RESULT 4
ICEC_MOUSE STANDARD: PRT; 419 AA.
ID ICEC_MOUSE
AC 008736;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Caspase-12 precursor (EC 3.4.22.-).
GN CASP12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C3H/An;
RX MEDLINE=97190206; PubMed=9038361;
RA van de Graen M., Vandenabeele P., Declercq W., van den Brande I.,
RA van Leeuwen G., Mollema F., Schotte P., van Ginkel W., Beyaert R.,
RA Fiers W.;
RT "Characterization of seven murine caspase family members";
RL FEBS Lett. 403:61-69(1997).
CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC RESPONSIBLE FOR APOPTOSIS EXECUTION (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS (POTENTIAL).
CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN SKELETAL MUSCLE AND LUNG.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14.
CC -1- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----

DR EMBL: Y13090; CAA73532.1; -
DR HSSP: P29466; 1ICE.
DR MEROPS: C14.013; -
DR MGD: MG1:1312922; Casp12.
DR InterPro: IPR001315; CARD.
DR InterPro: IPR002398; ICE.
DR InterPro: IPR002138; ICE_P10.
DR InterPro: IPR001309; ICE_P20.
DR Pfam: PF00655; ICE_P10; 1.
DR Pfam: PF00656; ICE_P20; 1.
DR PRINTS: PR00376; TLBCENZME.
DR SMART: SM00114; CARD; 1.
DR SMART: SM00115; CASc; 1.
DR PROSITE: PS0209; CARD; 1.
DR PROSITE: PS01122; CASPASE_CYS; 1.
DR PROSITE: PS01121; CASPASE_HIS; 1.
DR PROSITE: PS0207; CASPASE_P10; 1.
DR PROSITE: PS0208; CASPASE_P20; 1.
KW Hydrolyase; Thiol protease; Apoptosis; Zymogen.
FT PROPEP 1 ? ?
FT CHAIN ? ? CASPASE-12 SUBUNIT 1.
FT PROPEP ? ? POTENTIAL.
FT CHAIN ? ? CASPASE-12 SUBUNIT 2.
FT DOMAIN ? 419
FT ACT_SITE 1 92
FT ACT_SITE 250 250 CARD.
FT ACT_SITE 298 298 BY SIMILARITY.
SQ SEQUENCE 419 AA; 47853 MW; B94BOFEDI6LICB40 CRC64;

Query Match 13.2%; Score 75.5; DB 1; Length 419;
Best Local Similarity 35.6%; Pred. No. 5;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDYVKECYNYVVDGEEL---DQEKFYVD--ENFTENYV--TDCEK--DAGNAGTG 94
DB 26 VFDDLYEK---NVNGDELKIGESASFLINKAENLVENLEKTDMDAGKITAGHANSQ 81
QY 95 DE-----SDEVD 101
DB 82 EQLSIQFSNDEDD 94

RESULT 5

PC01_CAEEL STANDARD; PRT; 491 AA.
AC P49595;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Probable protein phosphatase 2C P4269.1 (EC 3.1.3.16) (PP2C).
GN P4269.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
RX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Tatch A.; Waterston R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
CC phosphate.
CC -1- CORACOR: BINDS 2 MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PP2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL: U00051; AAA91358.1; -

DR HSSP: P35813; 1A60.
DR WormPep: F4269.1; CE07231.
DR InterPro: IPR001932; PP2C-1ike.
DR InterPro: IPR000222; PP2C.
DR Pfam: PF00481; PP2C; 1.
DR SMART: SM00331; PP2C_SIG; 1.
DR SMART: SM00332; PP2C; 1.
DR PROSITE: PS01032; PP2C; 1.
KW Hypothetical protein; Hydrolyase; Magnesium; Manganese.
FT METAL 37 37
FT METAL 38 38 MANGANESE 1 (BY SIMILARITY).
FT METAL 57 57 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 428 428 MANGANESE 2 (BY SIMILARITY).
FT METAL 477 477 MANGANESE 2 (BY SIMILARITY).
SQ SEQUENCE 491 AA; 53141 MW; F110D12E435306 CRC64;

Query Match 13.2%; Score 75.5; DB 1; Length 491;
Best Local Similarity 24.5%; Pred. No. 5.9;
Matches 23; Conservative 18; Mismatches 28; Indels 25; Gaps 4;

QY 10 YANKRYTDF--DKSFKSSDDECKKTCFKETCYVFEYVNECYNYVVDGEELDOEK 67
DB 67 FTSAPLPFLERKRFWEADVAECLQKAF-----VFDDFIRAE-----ESMKELK 112
QY 68 FVVDENFTENYITDCEGKADGNAAGTGDESDVD 101
DB 113 DIGDEG-----KPKKAGEADSEDAERID 137

RESULT 6

YOFA_CAEEL STANDARD; PRT; 360 AA.

ID YOFA_CAEEL
AC Q09337;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 39.7 kDa protein ZK1290.10 in chromosome II precursor.
GN ZK1290.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
RX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Tatch A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL: U21308; AAB93316.1; -
DR WormPep: ZK1290.10; CE02080.
DR InterPro: IPR003582; SHKT.
DR SMART: SM00254; SHKT; 1.
KW Hypothetical protein; Signal.
FT SIGNAL 1 21
FT CHAIN 22 360
SQ SEQUENCE 360 AA; 39669 MW; 18751B6C062DAF7 CRC64;

Query Match 13.1%; Score 75; DB 1; Length 360;
Best Local Similarity 33.3%; Pred. No. 4.8;
Matches 23; Conservative 9; Mismatches 23; Indels 14; Gaps 3;

QY 46 EDYVKECYNYVVDGEELDEKFPV-----DENFTENYITDCEGKADGNA-----AAGTG 94
DB 224 EDYVGE---NVFDEKLPFKKIYIRNDGSTTENTTEQSTTEKTGTGKGTIEVSVELGEG 280

```

QY      95 DESDEVDED 103
      1 : : : : :
Db      281 DDEDEDD 289

RESULT 7
YJ23_AQUAE
ID      YJ23_AQUAE      STANDARD:      PRT;      296 AA.
AC      067752;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DE      Hypothetical protein AQ_1923 precursor.
GN      AQ_1923
OS      Aquifex aeolicus.
OC      Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC      Aquifex.
OX      NCB1_TaxID=63363;
RN      11
RP      SEQUENCE FROM N.A.
RC      STRAIN=VF5;
RX      MEDLINE=98196666; PubMed=9537320;
RA      Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA      Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA      Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT      "The complete genome of the hyperthermophilic bacterium Aquifex
RT      aeolicus.";
RL      Nature 392:353-358(1998).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
CC      EMBL: AE000763; AAC07719.1;
KW      Hypothetical protein; Signal; Complete proteome.
FT      SIGNAL 1 20 POTENTIAL.
FT      CHAIN 21 296 HYPOTHETICAL PROTEIN AQ_1923.
SQ      SEQUENCE 296 AA; 34566 MW; 9C88F4456AA9449C CRC64;

Query Match 12.9%; Score 74; DB 1; Length 296;
Best Local Similarity 24.2%; Pred. No. 4.8;
Matches 22; Conservative 18; Mismatches 21; Indels 30; Gaps 4;

QY      9 FVANRKYTD-----FDKSFKSSDLDCKKCTGKTEYCIYVE-----DTVNRECY 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      18 YFADVKYMDMKAKEQYEFYLDKTLRKSG-----KSKYKVDYKPIGSGIIIKDY 67
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      55 YNVVDEEELDQEKVYDENFENNYLTDCGK 85
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      68 YDRNGEKEEKEIEIIIEK-----LSETEGK 92
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
DYHC_ANTCR
ID      DYHC_ANTCR      STANDARD:      PRT;      4466 AA.
AC      P39057;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Dyein beta chain, ciliary.
OS      Anthocidaris crassispina (Sea urchin).
OC      Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC      Echinoidea; Echinoidea; Echinacea; Echinoida; Echinometridae;
OC      Anthocidaris.
OX      NCB1_TaxID=7629;
RN      11
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91326104; PubMed=1830928;
RA      Ogawa K.;

```

```

RT "Four ATP-binding sites in the midregion of the beta heavy chain of
RL dynein."
RL Nature 352:643-645(1991).
CC -1- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
CC FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC DYNEIN HAS ATPASE ACTIVITY.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA),
CC THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D01021; BAA00827.1; -.
CC DR PIR; S17231; S17231.
CC DR InterPro; IPR004273; Dynein_heavy.
CC DR Pfam; PF03028; Dynein_heavy; 1.
CC DR Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
CC Colled coil.
CC KW DOMAIN 482 502 COILED COIL (POTENTIAL).
CC FT DOMAIN 627 643 COILED COIL (POTENTIAL).
CC FT DOMAIN 734 805 COILED COIL (POTENTIAL).
CC FT DOMAIN 1036 1056 COILED COIL (POTENTIAL).
CC FT DOMAIN 1306 1337 COILED COIL (POTENTIAL).
CC FT DOMAIN 1443 1468 COILED COIL (POTENTIAL).
CC FT DOMAIN 1550 1978 MICROTUBULE-BINDING (POTENTIAL).
CC FT DOMAIN 3033 3134 COILED COIL (POTENTIAL).
CC FT DOMAIN 3263 3325 COILED COIL (POTENTIAL).
CC FT DOMAIN 3573 3642 COILED COIL (POTENTIAL).
CC FT NP_BIND 154 161 ATP (POTENTIAL).
CC FT NP_BIND 1852 1859 ATP (POTENTIAL).
CC FT NP_BIND 2133 2140 ATP (POTENTIAL).
CC FT NP_BIND 2460 2467 ATP (POTENTIAL).
CC FT NP_BIND 2805 2812 ATP (POTENTIAL).
CC SQ SEQUENCE 4466 AA; 511772 MW; C465CC5C6CD325D CRC64;
CC -----
Query Match 12.9%; Score 74; DB 1; Length 4466;
Best Local Similarity 25.9%; Pred. No. 81;
Matches 28; Conservative 14; Mismatches 28; Indels 38; Gaps 6;
OY 1 EEREDCWTFYANRKY-----TFPDKSFKSSDLE-----CR-----KTC 35
DB 460 EEFQECARVETFRPYDGLDPTCOEFLDEYEEFKYVFDDRLGSLICGFDCCGLEAA 519
OY 36 FKEVXY-----IVFEDTVKCEKYVYNNV---DGEILLDOEKFFVDEN 73
DB 520 FKMALDCTGFLDPRVYIRNDF---ECKYPIVLMYLDQELDQSKSEIYDEN 564
CC -----
RESULT 9
DPOA_SCHPO
ID DPOA_SCHPO STANDARD; PRT; 1405 AA.
AC P28040;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase alpha catalytic subunit (EC 2.7.7.7) (DNA polymerase
DE 1).
GN POLI OR SM7 OR SPAC35.06c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBL_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RM MEDLINE=91238692; PubMed=2034212;
RA Damagnez V., Tillitt J., de Recondo A.-M., Baldacci G.;

```



```

RT  "The POLI gene from the fission yeast, Schizosaccharomyces pombe,
RT  shows conserved amino acid blocks specific for eukaryotic DNA
RT  polymerases alpha."
RL  Mol. Gen. Genet. 226:182-189(1991).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=93140876; PubMed=8423854;
RA  Jagmohan S., Kiar A.J.S.;
RT  "DNA polymerase-alpha is essential for mating-type switching in
RT  fission yeast."
RL  Nature 361:271-273(1993).
RN  [3]
RP  SEQUENCE FROM N.A.
RX  STRAIN=972;
RX  MEDLINE=21848401; PubMed=11859360;
RA  Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA  Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA  Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA  Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA  Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA  Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA  James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA  Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA  Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA  Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA  Skelton J., Simmonds M., Squares R., Stevens S.,
RA  Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
RA  Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA  Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA  Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA  Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA  Beger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA  Goffeau A., Cadieu E., Dreano S., Gloux S., Leleaux V., Mottier S.,
RA  Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA  Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Rhode G.,
RA  Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA  Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA  Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA  Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RT  "The genome sequence of Schizosaccharomyces pombe.";
RL  Nature 413:871-880(2002).
CC  -1- FUNCTION: POLYMERASE ALPHA IN A COMPLEX WITH DNA PRIMASE IS A
CC  REPLICATIVE POLYMERASE.
CC  -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC  + [DNA](N).
CC  -1- SUBCELLULAR LOCATION: Nuclear
CC  -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC  ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC  DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC  -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: X58299; CAA41232.1; -
DR  EMBL: X69673; CAB57881.1; -
DR  EMBL: 299296; CAB16598.1; -
DR  PIR: S15993; DJZPA.
DR  InterPro: IPR002064; DNA_pol_B.
DR  InterPro: IPR004578; POL2.
DR  Pfam: PF00136; DNA_pol_B; 1.
DR  Pfam: PF03104; DNA_pol_B_exo; 1.
DR  PRINTS: PR00106; DNAPOLB.
DR  SMART: SM00486; POLBc; 1.
DR  TIGRPFAM: TIGR00592; pol2; 1.
DR  PROSITE: PS00116; DNA_POLYMERASE_B; 1.
KW  Transferase; DNA-directed DNA polymerase; DNA replication;
KW  DNA-binding; Nuclear protein.

```

```

SQ  SEQUENCE 1405 AA; 159348 MW; 06082A3BAD6347C2 CRC64;
Query Match
Best Local Similarity 32.3%; Score 73; DB 1; Length 1405;
Matches 21; Conservative 10; Mismatches 18; Indels 16; Gaps 3;
QY 45 FEDATNKECYNNVNGE-----ELDEKRVNDEN--FENYITDCEGDAGNAAGTG 94
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 31 YESKNEELYNVSEEEYRKIVRQLRDLDDDFVDDNGAGYVDN-----GYDEWDQSHYS 84
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 95 DESDE 99
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 85 DEDDE 89
RESULT 10
YVW2_YEAST STANDARD; PRT; 782 AA.
ID YVW2_YEAST
AC PA0157;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 88.8 kDa protein in PEX17-MER1 intergenic region.
GN YNL212W OR N1327.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / FY1679;
RX MEDLINE=95282517; PubMed=7762305;
RA Coster F., van Dyck L., Jonniaux J.-L., Purnelle B., Goffeau A.;
RT "The sequence of a 13.5 kb DNA segment from the left arm of yeast
RT chromosome XIV reveals MER1, RAP1, a new putative member of the DNA
RT replication complex and a new putative serine/threonine phosphatase
RT gene."
RL Yeast 11:85-91(1995).
CC -1- SIMILARITY: TO C.CARDUNCULUS PROTEIN CYP04 AND S.POMBE
CC SPEC1685.14C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X78898; CAA55496.1; -
DR EMBL: Z71488; CAA96114.1; -
DR SGD: S0005156; YNL212W.
KW Hypothetical protein.
SQ SEQUENCE 782 AA; 88845 MW; CCBBI4FCE32BF97C CRC64;
Query Match
Best Local Similarity 26.5%; Score 71; DB 1; Length 782;
Matches 22; Conservative 16; Mismatches 39; Indels 6; Gaps 2;
QY 24 KSSDDECKKTCFTEYCIYFEDTVNKECYNNVNDSE--LDQKRVVDENFTENVLT 80
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 315 KFKDLDKICIQRFAMVNCIMW---TLNKEITVDPEREKVDIIIDSSVPLEKQDDLIHI 371
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 81 DCEGKDAGNAAGTDESDEVD 103
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 372 DDRSNERDKESSESENDESD 394
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
RESULT 11
MOH1_CAMPG STANDARD; PRT; 78 AA.
ID MOH1_CAMPG
AC P81034;
DT 15-JUL-1998 (Rel. 36, Created)

```

DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Mandibular organ-inhibiting hormone-1 (MOIH-1).
 OS Cancer pagurus (Crab crab).
 CC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
 CC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata;
 CC Brachyura; Eubrachyura; Cancroidea; Cancridae; Cancer.
 OX NCBI_TaxID=6755;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Stinus gland;
 RX MEDLINE=9627865; Pubmed=8662685;
 RA Malinwight G., Webster S.G., Wilkinson M.C., Chung J.S., Rees H.H.;
 RT "Structure and significance of mandibular organ-inhibiting hormone in
 the crab, Cancer pagurus. Involvement in multihormonal regulation of
 growth and reproduction.";
 RL J. Biol. Chem. 271:12749-12754(1996).
 CC -1- FUNCTION: REPRESENTS THE SYNTHESIS OF METHYL FARNESATE, THE
 PRECURSOR OF INSECT JUVENILE HORMONE III IN THE MANDIBULAR ORGAN.
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE MEDULLA TERMINALIS X-ORGAN IN
 THE EYES/TALKS AND TRANSPORTED TO THE SINUS GLAND WHERE IT IS
 STORED AND RELEASED.
 CC -1- SIMILARITY: BELONGS TO THE ARTHROPOD CHH/MIH/GIH/VIH FAMILY OF
 HORMONES.
 CC InterPro: IPR001166; CHH_MIH_GIH.
 DR Pfam: PF01147; Crust_neurohorm; 1.
 DR PRINTS: PR00550; HYPRGLYCEMIC.
 DR PROSITE: PS01250; CHH_MIH_GIH, 1.
 KW Neuropeptide; Hormone
 FT DISULFID 7 44 BY SIMILARITY.
 FT DISULFID 24 40 BY SIMILARITY.
 FT DISULFID 27 53 BY SIMILARITY.
 SQ SEQUENCE 78 AA; 9241 MW; 8DB338A39058A62D CRC64;

Query Match 12.3%; Score 70.5; DB 1; Length 78;
 Best Local Similarity 33.9%; Pred. No. 2.5;
 Matches 20; Conservative 5; Mismatches 23; Indels 11; Gaps 3;

OY 5 DCMWTFYANKRYK-----TDFDKSPFKSSDLDCKKTCF-KTEY--CYVFEDYVKE 52
 DB 6 DQNMIGRAMYKVEYDMCKNCANIFRKDGLNMCNSCYNTEFLMCIDATENTRKE 64

RESULT 12
 PA21_CANFA STANDARD; PRT; 146 AA.
 AC P06596;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
 acylhydrolase) (Group IB phospholipase A2).
 GN PLA2G1B.
 OS Canis familiaris (Dog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=66223862; Pubmed=3754861;
 RA Ohara O., Tamaki M., Nakamura E., Tsuruta Y., Fujii Y., Shln M.,
 RT "Dog and rat pancreatic phospholipases A2: complete amino acid
 sequences deduced from complementary DNAs.";
 RL J. Biochem. 99:733-739(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90306027; Pubmed=2142076;
 RA Kefelef B., Laforge K.S., Vasiloudes P., Puigserver A., Scheele G.A.;
 RT "Isolation and sequence of the canine pancreatic phospholipase A2
 gene.";
 RL Eur. J. Biochem. 190:299-304(1990).
 RN [3]

RP SEQUENCE FROM N.A.
 RX MEDLINE=87175472; Pubmed=3562437;
 RA Kefelef B., Laforge K.S., Puigserver A., Scheele G.A.;
 RT "Primary structures of canine pancreatic lipase and phospholipase A2
 messenger RNAs.";
 RL Pancreas 1:430-437(1986).
 CC -1- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the
 2-acyl groups in 3-sn-phosphoglycerides.
 CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O - 1-
 acylglycerophosphocholine + a fatty acid anion.
 CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D00035; BAA00023.1; -;
 DR EMBL: M35301; AAA30883.1; -;
 DR PIR: A24392; PSDG.
 DR PIR: S11316; S11316.
 DR HSP: P00592; 4P2P.
 DR InterPro: IPR001211; PhospholipaseA2.
 DR Pfam: PF000068; phosph1p; 1.
 DR PRINTS: PR00389; PHPLIPASEA2.
 DR ProDom: PD000303; PhospholipaseA2.
 DR SMART: SM00085; PA2c; 1.
 DR PROSITE: PS00118; PA2_HIS; 1.
 DR PROSITE: PS00119; PA2_ASP; 1.
 KW Hydrolase; Lipid degradation; Calcium; Pancreas; Signal.
 FT SIGNAL 1 15
 FT PROPEP 16 22
 FT CHAIN 23 146 PHOSPHOLIPASE A2.
 FT ACT_SITE 70 70 BY SIMILARITY.
 FT ACT_SITE 121 121 BY SIMILARITY.
 FT DISULFID 33 99 BY SIMILARITY.
 FT DISULFID 49 146 BY SIMILARITY.
 FT DISULFID 51 67 BY SIMILARITY.
 FT DISULFID 66 127 BY SIMILARITY.
 FT DISULFID 73 120 BY SIMILARITY.
 FT DISULFID 83 113 BY SIMILARITY.
 FT DISULFID 106 118 BY SIMILARITY.
 FT CA_BIND 50 50 VIA CARBONYL OXYGEN (BY SIMILARITY).
 FT CA_BIND 52 52 VIA CARBONYL OXYGEN (BY SIMILARITY).
 FT CA_BIND 54 54 VIA CARBONYL OXYGEN (BY SIMILARITY).
 FT CA_BIND 71 71 BY SIMILARITY.
 SQ SEQUENCE 146 AA; 16235 MW; F6258ED9527F3692 CRC64;

Query Match 12.3%; Score 70.5; DB 1; Length 146;
 Best Local Similarity 25.4%; Pred. No. 4.9;
 Matches 15; Conservative 11; Mismatches 18; Indels 15; Gaps 1;

OY 28 LDPCCKTKCFKTECYVFEDYVANKECYVNWVDEELDQKFPVDENFENYLTDEGKD 86
 DB 60 VDELDKCCQTHDHCY-----SEAKILDSCKFLDNPYRKYISYSCSGSE 103

RESULT 13
 CB20_HUMAN STANDARD; PRT; 156 AA.
 AC P52298; Q14924;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 20 kDa nuclear cap binding protein (NCBP 20 kDa subunit) (CBP20) (NCBP
 interacting protein 1) (NIP1).
 GN NCBP2 OR CBP20.
 OS Homo sapiens (Human).
 RN [3]

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBITaxid=9606;
 RN [1]
 RX MEDLINE=95379956; PubMed=7651522;
 RA Izaurralde E., Lewis J., Gamberi C., Jarmolowski A.,
 RA McGuigan C., Malta J.A.W.,
 RT "A cap binding protein complex mediating U snRNA export.";
 RL Nature 376:709-712(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=96017765; PubMed=7478990;
 RA Kataoka N., Ohno M., Mada I., Shimura Y.,
 RT "Identification of the factors that interact with NCBP, an 80 kDa
 RT nuclear cap binding protein.";
 RL Nucleic Acids Res. 23:3638-3641(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cervix;
 RA Strausberg R.,
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 FUNCTION: INVOLVED IN MEDIATING U SNRNA EXPORT FROM THE NUCLEUS.
 CC BINDS TO 5'CAPED MRNA.
 CC -1 SUBUNIT: THE CAP-BINDING PROTEIN (CBC) COMPLEX IS AN HETERODIMER
 CC OF CBP80 AND CBP20.
 CC -1 SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X84157; CAAS8962.1; -
 DR EMBL: D59253; BAA09599.1; -
 DR EMBL: BC001255; AAH01255.1; -
 DR HSSP: P11940; IGVJ.
 DR Genew: HGNC:7659; NCBP2.
 DR MIM: 605133; -
 DR InterPro: IPR005054; RNA_rec_mot.
 DR Pfam: PFO0076; rrm.1.
 DR SMART: SMO0360; RRM.1.
 DR PROSITE: PS00103; RRM.1.
 DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
 DR Nuclear protein; RNA-binding.
 KW DOMAIN 40 118 RNA-BINDING (RRM).
 FT CONFLICT 97 97 A -> S (IN REF. 2).
 FT SEQUENCE 156 AA; 18001 MW; B6C94F3182A2CC3D CRC64;
 SQ
 Query Match 12.3%; Score 70.5; DB 1; Length 156;
 Best Local Similarity 21.8%; Pred. No. 5.2;
 Matches 26; Conservative 20; Mismatches 50; Indels 23; Gaps 6;
 QY 1 EERE-----DCWTFYANKRYTDFKS-----FKKSSD-----LDDECKTKCFTEYCYIV 44
 DB 30 EDEGKLKRSCTLYGVNLSPTTTEQIYELFSSGDKTKITIMGLDKKTKTA--CGFCFE 87
 QY 45 FEDTVNEKCYVNVVGEELDEKFEVDENFTENTYLTDEGKDGAGNAAGTDESDVEVD 103
 DB 88 YSRADAEAMRYINGTRILDDR--IIRTDMAGFK---EGRQYGRGRSGGVADERYQD 141
 RESULT 14
 Y875_METJA STANDARD; PRT; 748 AA.
 AC 058285;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0875.
 GN MJ0875.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 NC NCBITaxid=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerecavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Ustebach T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii.";
 RL Science 273:1058-1073(1996).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U67531; AAB98882.1; -
 DR TIGR: MJ0875; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 748 AA; 90731 MW; 515ED3B541562EA CRC64;
 SQ
 Query Match 12.3%; Score 70.5; DB 1; Length 748;
 Best Local Similarity 24.4%; Pred. No. 27;
 Matches 29; Conservative 16; Mismatches 39; Indels 35; Gaps 6;
 QY 1 EEREDCWTFY-----ANKRYTDFKSFKSSDLDECKTKCFKTEYCYIVFDD 47
 DB 144 KEFSDIWMYFEDIKYYKKLSIKHRKEGDLKAIETLTK---KCYLAESYVKKFNDK 200
 QY 48 ----TYNKECYVNVVGEELDEKFEVDENFTENTYLTDEGKDGAGNAAGTDESDVEVD 102
 DB 201 KKAEIFNKHRYNLM-----AKFESERKFE-----AAEYKKRSGDTIKIDE 244
 RESULT 15
 S185_YEAST STANDARD; PRT; 1058 AA.
 AC P40856;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE S174-associating protein SAPI85.
 GN SAPI85 OR YJ1098W OR J0840.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NC NCBITaxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96220458; PubMed=8649382;
 RA Luke M.M., della Seta F., di Como C.J., Sugimoto H., Kobayashi R.,
 RA Arndt K.T.;
 RT "The SAP, a new family of proteins, associate and function positively
 RT with the S174 phosphatase.";
 RL Mol. Cell. Biol. 16:2744-2755(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;

1	78.5	13.7	116	16	08XJ41	08XJ41 clostridium
2	78.5	13.7	509	2	09X9S1	09X9S1 streptococc
3	78.5	13.7	550	5	09U2D4	09U2D4 caenorhabdi
4	78.5	13.7	907	5	09XTS9	09XTS9 caenorhabdi
5	78	13.6	682	16	09K6N5	09K6N5 bacillus ha
6	78	13.6	3933	5	09J239	09J239 plasmodium
7	77	13.4	368	5	09F7H8	09F7H8 drosophila
8	77	13.4	374	5	08SYG5	08SYG5 drosophila
9	77	13.4	945	5	003650	003650 trypanosoma
10	75.5	13.2	264	5	020246	020246 caenorhabdi
11	75	13.1	3844	5	094648	094648 plasmodium
12	74.5	13.0	606	10	09FTW8	09FTW8 arabidopsis
13	74.5	13.0	1277	11	035821	035821 rattus norv
14	74	12.9	596	10	09FK41	09FK41 arabidopsis
15	74	12.9	1278	16	050667	050667 borrelia bu
16	73.5	12.8	156	11	09CQ49	09CQ49 mus musculus

17	73.5	12.8	1154	5	Q9V6S9	Q9V6S9 dtrosophila
18	73	12.7	643	5	Q9XV53	Q9XV53 caenorhabd1
19	73	12.7	2543	5	P80649	P80649 parametium
20	72.5	12.7	564	5	O01627	O01627 caenorhabd1
21	72.5	12.7	2533	5	Q27183	Q27183 parametium
22	72.5	12.7	2533	5	P90589	P90589 parametium
23	72	12.6	1303	10	Q9SKS0	Q9SKS0 arabidopsist
24	71.5	12.5	320	5	O96214	O96214 plasmedium
25	71.5	12.5	540	16	O8XHG9	O8XHG9 clostridium
26	71.5	12.5	594	16	O9PPE1	O9PPE1 campylobact
27	71	12.4	551	4	Q9NX93	Q9NX93 homo sapien
28	71	12.4	589	4	Q9HA35	Q9HA35 homo sapien
29	71	12.4	851	4	Q9H501	Q9H501 homo sapien
30	70.5	12.3	112	5	Q9U5Y6	Q9U5Y6 cancer pagu
31	70.5	12.3	256	10	Q9CAs9	Q9CAs9 arabidopsist
32	70.5	12.3	296	11	O9DS56	O9DS56 mus musculu1
33	70.5	12.3	367	5	Q9V1R8	Q9V1R8 dtrosophila
34	70.5	12.3	600	16	O8RHS6	O8RHS6 fusobacter
35	70.5	12.3	786	10	Q9C7E7	Q9C7E7 arabidopsist
36	70.5	12.3	2472	5	O8T2M5	O8T2M5 dictyostell1
37	70	12.2	292	12	O65239	O65239 african sw
38	70	12.2	328	5	O22144	O22144 caenorhabd1
39	70	12.2	521	5	O94718	O94718 parametium
40	70	12.2	669	5	O8XS58	O8XS58 dtrosophila
41	70	12.2	793	10	Q9LKR9	Q9LKR9 arabidopsist
42	70	12.2	793	10	Q9SPU5	Q9SPU5 zea mays (m
43	70	12.2	1167	16	O9ZLG7	O9ZLG7 helicobacter
44	70	12.2	1255	16	O971T6	O971T6 clostridium
45	70	12.2	1481	5	Q9NEX0	Q9NEX0 caenorhabd1

RESULT 1

ID	Q8XJ41	PRELIMINARY;	PRT;	116 AA
----	--------	--------------	------	--------

DT 01-MAR-2002 (TREMBLÉ, 20, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

GN CPE1920.

OC Bacteria; Firmicutes; Bacillu

OX NCBI_TaxID=1502;

RP SEQUENCE FROM N.A.

RX PubMed=11792842;

RA Shiba T., Ogasawara N., Hattori T.

```
RT flesh-eater.";
```

DR EMBL; AP003192; BAB81626.1; -

DR Pfam; PF00515; TPR; 3.

KW Hypothetical protein; Complete

Best Local Similarity 22.78;

—
—
—
—
• •
• •
—
• •
—
• •

Db 71 ENLAYCYSMINNSKALRYFNALMDNADIDCEKAI 107

RESULT 2

09X9S1 PRELIMINARY; PRT; 509 AA.

AC 09X9S1; 1

DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Polysaccharide synthase.

GN TTS.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Streptococcus.

OX NCBI_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RA Llull D., Munoz R., Lopez R., Garcia E.;

RT "A single gene (tts) located outside the cap locus directs the formation of Streptococcus pneumoniae type 37 capsular polysaccharide: RT type 37 pneumococci are natural, genetically binary strains.";

RL J. Exp. Med. 19:241-251(1999).

DR EMBL: AJ131985; CAB51329.1;

DR InterPro: IPR001173; Glycos_transf_2.

DR Pfam: PF00535; Glycos_transf_2; 1.

SO SEQUENCE 509 AA; 58888 MW; C6C38D2ACFAB0D8A CRC64;

Query Match 13.7%; Score 78.5; DB 2; Length 509;

Best Local Similarity 31.1%; Pred. No. 9.8;

Matches 23; Conservative 10; Mismatches 30; Indels 11; Gaps 3;

OY 8 TTYANRKYTDFDKSEKSSDDECKKTCFKTECYIVFEDTVN---KECYNVVDGEELD 64

DB 74 TTYLNNYIHIGSSPKKSSISEAKKVILL--YC-----TANFVECLVESMQDYAN 125

OY 65 OEKPVVDENFTENY 78

DB 126 FETVILDDSKSEVY 139

RESULT 3

09U2Q4 PRELIMINARY; PRT; 550 AA.

AC 09U2Q4; 1

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE Y26D4A.2 protein.

GN Y26D4A.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA White S.;

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for investigating biology.";

RL Science 282:2012-2018(1998).

DR EMBL: AL110478; CAB54340.1; -

DR InterPro: IPR001304; Lectin_C.

DR SMART: SM00034; CLECT; 1.

SO SEQUENCE 550 AA; 62195 MW; 79AE220BE127466A CRC64;

Query Match 13.7%; Score 78.5; DB 5; Length 550;

Best Local Similarity 25.6%; Pred. No. 11;

Matches 23; Conservative 12; Mismatches 32; Indels 23; Gaps 3;

OY 6 CWTYANRKYTDF-----DKSEKKS--SDDECKKTCFKTECYI 43

DB 170 CGNFLLNRPYPDSCETECMDTWQVNAKPGPLMGSSNINDNLNLTSSDECYKCYCMKDLNCFV 229

OY 44 VFEDTVNKECYNVVDGEELDEKFEVDEN 73

DB 230 NYDKDSKRCQWMSIDNVHF-LEKVHPSN 258

RESULT 4

09XTS9 PRELIMINARY; PRT; 907 AA.

AC 09XTS9; 1

DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Y69H2.2 protein.

GN Y69H2.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA McMurray A.A.;

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for investigating biology.";

RL Science 282:2012-2018(1998).

DR EMBL: Z98877; CAB11570.1; -

DR HSP: Q12784; 1HSP.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001458; EGF-11.

DR InterPro: IPR001873; Na+channel_ASC.

DR Pfam: PF00858; ASC; 1.

DR Pfam: PF00008; EGF; 7.

DR PRINTS: PRO1078; AMINACHANNEL.

DR PRINTS: PRO0010; EGFLOOD.

DR SMART: SM00181; EGF; 8.

DR PROSITE: PS00022; EGF_1; UNKNOWN. 8.

DR PROSITE: PS01186; EGF_2; 4.

KW EGF-like domain; Glycoprotein.

SO SEQUENCE 907 AA; 101840 MW; 8B55FF139F8F81D7 CRC64;

Query Match 13.7%; Score 78.5; DB 5; Length 907;

Best Local Similarity 29.2%; Pred. No. 18;

Matches 33; Conservative 12; Mismatches 33; Indels 35; Gaps 8;

OY 10 YANRKYTDFDKSKF-----KSSD-----DECKKTC--FKTECYIVFEDT 48

DB 226 YAOEKSTD-DKTKLYTLGECWCVSEDMVKDDVSMKDCSGGDSGYC-----GK 279

OY 49 VNKECYNVVDGEELDEKFEVDENFTENYLVNDCESKAGNAAGTGESDEVD 101

DB 280 VNSRCI--VYENGEME-----TDPNACEDNSTLC-GAELGHGCMIMQSDVD 324

RESULT 5

09K8N5 PRELIMINARY; PRT; 682 AA.

AC 09K8N5; 1

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Two-component sensor histidine kinase involved in chemotaxis.

GN CHEA OR BH2970.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Bacillaceae; Bacillus.
 RN NCB1_TaxID=86665;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 CC -1- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
 DR EMBL: AP001517; BAB06689.1; -
 DR HSSP: 056310; 1830.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR004358; Bact_sens_pr_C.
 DR InterPro: IPR002545; Chem.
 DR InterPro: IPR004105; H-Kinase_dlm.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR InterPro: IPR002570; Hpt.
 DR Pfam: PF01584; Chem; 1.
 DR Pfam: PF02895; H-Kinase_dlm; 1.
 DR Pfam: PF02517; HATPase_C; 1.
 DR Pfam: PF01627; Hpt; 1.
 DR PRINTS: PR00344; BCTRLENSOR.
 DR PRODOM: PD003142; Hpt; 1.
 DR SMART: SM00260; Chem; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00073; HPT; 1.
 DR Kase: Phosphorylation; Sensory transduction; Transferase;
 KW Complete proteome.
 SQ SEQUENCE 682 AA; 75272 MW; 46B61BEE5BF7276 CRC64;

Query Match 13.6%; Score 78; DB 16; Length 682;
 Best Local Similarity 32.6%; Pred. No. 15;
 Matches 28; Conservative 13; Mismatches 31; Indels 14; Gaps 5;

OY 10 YANKRYTDPDKRFFKSSD-----LDECKTKCFTECYIYFE--DYVNECYIYVVDG 60
 DB 156 YDEFTETVLDOSFEQGYQAYQLEVTLD--KTLKAAEFVMEVLEQVG--EVIKSPSA 212
 OY 61 EELDOKEFYVDENFENTYLTDGDC 86
 DB 213 EELDEKFF--DERFLVTLTKVDGE 236
 RESULT 6
 ID 097239 PRELIMINARY; PRT; 3933 AA.
 AC 097239;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 467.9 kDa protein.
 GN PFC0245C, MAL3P2.18.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCB1_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=99376085; PubMed=10448855;
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltham T.,
 RA Genies S., Gilliam R., Hamlin N., Harris R., Holtroyd S., Hornsby T.,
 RA Hurrell K., Jagsall K., Jassal B., Kyes S., McLean J., Moule S.,
 RA Murrell K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
 RA Rutter S., Skellern J., Squares R., Squares S., Sultson J.E.,
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.,
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
 RT falciparum";
 RL Nature 400:532-538(1999).

DR EMBL: AL034558; CAB39005.1; -
 DR InterPro: IPR002048; EF-hand.
 DR ProSITE: PS00018; EF_HAND; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 3933 AA; 467876 MW; 51444604EE36933 CRC64;

Query Match 13.6%; Score 78; DB 5; Length 3933;
 Best Local Similarity 25.0%; Pred. No. 1e+02;
 Matches 23; Conservative 21; Mismatches 24; Indels 24; Gaps 5;

OY 36 FKTEYC-YIYF-----EDTVNKECY---NVVDGEELD-QEKF-----VVD 71
 DB 1139 FFFNNCEYIIFLYNNHLLIEKEKLNKSCYFLKLNLTNCTCNKKNKFFWCFELPHIIR 1198
 OY 72 ENFENYLTDEGKADGNAAGTDESEVED 103
 DB 1199 INFKSLKLNKIKKEAGDDTDDDDDDDDDD 1230

RESULT 7
 ID 09VFH8 PRELIMINARY; PRT; 368 AA.
 AC 09VFH8;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG14853 protein.
 GN CG14853.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCB1_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman M.D., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
 RA April J.E., Abghyani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Boucek J., Brokstein P., Brottier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferrara C., Ferrieri S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jalaia M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kelchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Moshrefi A.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Polard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner A., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasgaman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;


```
Query Match      13.2%   Score 75.5; DB 5; Length 264;  
Best Local Similarity 27.7%; Pred. No. 9.4;  
Matches    31; Conservative 19; Mismatches 33; Indels 29; Gaps 8;
```

OY 1 EEEEDCWTFYANKRYTDEDFKSFKK--SDDLDECCKTCETEXCYATFE-----DTYNKE 52
 ||| | :|| | :|| |
Db 29 EESEED-----FEKVFLAIIISKHCESK----DYDFARLIQRREALINSGKE 72

OY 53 CYNVVDGEBLQDKF-VYDENFTF--NYLLDCGGKAAGNAGCGDSDDVD 101
 |: :| | |:::||:::| ::||: |::| :||: :

Db 73 CFEIIVK-EECPPEKFKLIEENYSQVLTLLTE-KPKDNGACTAPYFQLEEIE 122

RESULT 11

ID	Q94648	PRELIMINARY;	PRT;	3844	AA.
AC	Q94648.				

DT	01-FEB-1997	(TREMBLrel. 02, Created)
DT	01-JUL-1997	(TREMBLrel. 04, Last sequence update)
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)
DE	AAPI1 protein (Fragment).	
GN	AAPI1.	
OS	Plasmodium falciparum (isolate 3D7).	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_TaxID=36329;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=3D7;	
RX	MEDLINE=97378065; PubMed=9234746;	
RA	Barral J.C., Candelle D., Attal-Bonnefoy G., Dehoux P., Bonnefoy S.,	
RA	Ridley R., Pereira da Silva L., Lampsley G.;	
RT	"Plasmodium falciparum AAPI1, a giant protein containing repeated	
RT	motifs rich in asparagine and aspartate residues, is associated with	
RT	the infected erythrocyte membrane.";	
LT	Infect. Immun. 65:3013-3010(1997).	
DR	EMBL: Y08926; CAA70130.1; -;	
DR	InterPro: IPRO01841; znf_r1ng.	
DR	SMART: SM00184; R1NG.1.	
FT	NON_TER	1
FT	NON_TER	3844
SQ	SEQUENCE	3844 AA; 456075 MW; 23D0233261C30DBB CRC64;

Query Match	13.1%;	Score	75;	DB	5;	Length	3844;
Best Local Similarity	23.9%;	Pred. No.	2e+02;				
Matches	21;	Conservative	23;	Mismatches	34;	Indels	10;
						Gaps	3;

```

QY      1  EERDDCCTFAANKRYTD-----PDR-SFKSSDLDCKCTCKPTCYATVEDPYNK 51
      1726  EDSEQCIRKSYKKMSGCNNKENIIFDSTSVLRKNNIRLRKNYMKNNCNCIYYDDNNK 1785

Db
QY      52  ECYNYVDGELDOEKFYVDENTFYL 79
      1786  K-KNNKNVNDQREKYTLAKTIFVHNFI 1812

Db

```

RESULT 12	
Q9FIW8	
ID Q9FIW8	PRELIMINARY;
AC Q9FIW8:	PRT; 606 AA.

DT 01-MAR-2001 (TREMBLrel. 10, Creaseau)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE GB|AAE22924.1 (Hypothetical 71.0 kDa protein).
 GN AF5639780.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae
 OC eumossids II; Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_Taxid=3702;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=9087489; PubMed=9872454;
 RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
 RT Sequence features of the regions of 1,013,767 bp covered by sixteen
 RT physically assigned P1 and TAC clones."
 RL DNA Res. 5:297-308(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Bann J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Totluni M., Wu H.C.,

RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamita A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narasaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Souhewick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene At5g39780 (GI:15242471).";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016876; BAB11383.1; -
 DR EMBL; AY070102; AAL49839.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 606 AA; 71043 MW; 5F02E244E59FC15D CRC64;

Query Match 13.0%; Score 74.5; DB 10; Length 606;
 Best Local Similarity 27.7%; Pred. No. 29;
 Matches 31; Conservative 19; Mismatches 37; Indels 25; Gaps 7;

QY 1 EREEDCW--TFYANKRYTDFD--KSFKSSDLDECKTCFTECYIVFEDTVKCEY 55
 DB 103 EKTEDLGIVFHGEDKVKTEDYSVSSFK-----KKIRFLTE-----EDFLESDS-- 147
 QY 56 NVVDEEL----DQEFVVDENFTENYLTDCGKDGAGNAGTGDSEDEVED 103
 DB 148 DFVDSQGTFTSNDDEGFLSDSDFATSLKKGGRKSDN--SGSGSDSEEEEEE 198

RESULT 13
 ID 035821 PRELIMINARY; PRT; 1277 AA.
 AC 035821;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PAR interacting protein.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91084854; PubMed=2261643;
 RA Wharlin J., Schibler U.;
 RT "Expression of the liver-enriched transcriptional activator protein
 RT DBP follows a stringent circadian rhythm.";
 RL Cell 63:1257-1266(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Comte P.A., Ossipow V., Schibler U.;
 RT "Isolation of PIP, a 160 kDa nuclear protein that interacts with the
 RT activation domain of PAR transcription factors.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U83590; AAB62878.1; -
 SQ SEQUENCE 1277 AA; 144675 MW; 43140A2973176792 CRC64;

Query Match 13.0%; Score 74.5; DB 11; Length 1277;
 Best Local Similarity 30.4%; Pred. No. 66;
 Matches 31; Conservative 9; Mismatches 37; Indels 25; Gaps 5;

QY 1 EREEDCWTFYANKRYTDFD--KSFKSSDLDECKTCFTECYIVFEDTVKCEY 59
 DB 630 EDEED-----NVVYTDDEKOLKHEGADSDS-----EDSKSESVDSEED 670
 QY 60 GEELDOE--KVVVDENFTENYLTDCGKDGAGNAGTGDSEDE 99
 DB 671 GESEEDRDKVDVDPGFRQOLM---EVLQAGNALGGESEEE 709

RESULT 14
 ID 09FK41 PRELIMINARY; PRT; 596 AA.
 AC 09FK41;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Similarity to unknown protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=98403884; PubMed=9734815;
 RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
 RT Sequence features of the regions of 1,367,185 bp covered by 19
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:203-216(1998).
 DR EMBL; AB012248; BAB09450.1; -
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF01344; Kelch; 6.
 SQ SEQUENCE 596 AA; 68483 MW; 8865D3CAD873CB0D CRC64;

Query Match 12.9%; Score 74; DB 10; Length 596;
 Best Local Similarity 24.2%; Pred. No. 32;
 Matches 29; Conservative 23; Mismatches 40; Indels 28; Gaps 5;

QY 12 NRKYTDFDKSFK-----KSSDLDECKTCFTECYIVFEDTV-----NKEC 53
 DB 339 NRRMSDKSKALQGAKTRLDPQVSSEVVPKCRINSNCVVGHDTLYGGMKMDREV 398
 QY 54 YYN---VVDGEELDQEFVVDENFT--NYLTDCG-----KAGNAGTGDSEDEVED 103
 DB 399 TLDDLYSLNLSKIDEMWCIPTTETEWVEVSDEGEDEDEDESEGESESEDEDD 458

RESULT 15
 ID 050667 PRELIMINARY; PRT; 1278 AA.
 AC 050667;
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Hypothetical protein BBH09.
 GN BBH09.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Gasjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Uterback T., Matthey L., McDonald L., Artiach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 RT burgdorferi.";
 RL Nature 390:580-586(1997).
 DR EMBL; AE000784; AAC66000.1; -
 DR TIGR; BBH09; -
 DR InterPro; IPR002296; N12N6_mtfase.
 DR InterPro; IPR002052; N6_mtfase.
 DR PRINTS; PR00507; N12N6MTFASE.
 DR PROSITE; PS00092; N6_MTFASE; UNKNOWN_1.
 KW Hypothetical protein; Pfam; Complete proteome.
 SQ SEQUENCE 1278 AA; 150871 MW; 8A0B22DF16565C0 CRC64;

Query Match 12.9%; Score 74; DB 16; Length 1278;

Tue May 6 16:59:49 2003

us-09-936-737a-2.rspt

Page 7

Best Local Similarity 25.0%; Pred. No. 74;
Matches 18; Conservative 17; Mismatches 25; Indels 12; Gaps 3;

```
OY      10 YANKRYIDFDSFKKSDLBCKCTCKTXYCYVFEEDJVAKCEYNW-----VDGEILD 64  
        |::||| | : | :: ||| | | | |::|  
Db     181 YGCRKYIEFD--FSKIRKEE----YKEOWFLIFSLYLRKERYLTSMIISVEKKQIS 233
```

```
QY      65 QEFVVDENFTE 76
        :||:::|
Db      234 KEKEIQKTLRE 245
```

```
Search completed: May 5, 2003, 17:52:37
Job time : 71 secs
```

